Introduction to Science and the Realm of Physics, Physical Quantities, and Units

class="introduction"

Galaxies are as immense as atoms are small. Yet the same laws of physics describe both, and all the rest of nature—an indication of the underlying unity in the universe. The laws of physics are surprisingly few in number, implying an underlying simplicity to nature's apparent complexity. (credit: NASA, JPL-Caltech, P. Barmby, Harvard-Smithsonian Center for

Astrophysics
)

What is your first reaction when you hear the word "physics"? Did you imagine working through difficult equations or memorizing formulas that seem to have no real use in life outside the physics classroom? Many people come to the subject of physics with a bit of fear. But as you begin your exploration of this broad-ranging subject, you may soon come to realize that physics plays a much larger role in your life than you first thought, no matter your life goals or career choice.

For example, take a look at the image above. This image is of the Andromeda Galaxy, which contains billions of individual stars, huge clouds of gas, and dust. Two smaller galaxies are also visible as bright blue spots in the background. At a staggering 2.5 million light years from the Earth, this galaxy is the nearest one to our own galaxy (which is called the Milky Way). The stars and planets that make up Andromeda might seem to be the furthest thing from most people's regular, everyday lives. But Andromeda is a great starting point to think about the forces that hold together the universe. The forces that cause Andromeda to act as it does are the same forces we contend with here on Earth, whether we are planning to send a rocket into space or simply raise the walls for a new home. The same gravity that causes the stars of Andromeda to rotate and revolve also causes water to flow over hydroelectric dams here on Earth. Tonight, take a moment to look up at the stars. The forces out there are the same as the ones here on Earth. Through a study of physics, you may gain a greater

understanding of the interconnectedness of everything we can see and know in this universe.

Think now about all of the technological devices that you use on a regular basis. Computers, smart phones, GPS systems, MP3 players, and satellite radio might come to mind. Next, think about the most exciting modern technologies that you have heard about in the news, such as trains that levitate above tracks, “invisibility cloaks” that bend light around them, and microscopic robots that fight cancer cells in our bodies. All of these groundbreaking advancements, commonplace or unbelievable, rely on the principles of physics. Aside from playing a significant role in technology, professionals such as engineers, pilots, physicians, physical therapists, electricians, and computer programmers apply physics concepts in their daily work. For example, a pilot must understand how wind forces affect a flight path and a physical therapist must understand how the muscles in the body experience forces as they move and bend. As you will learn in this text, physics principles are propelling new, exciting technologies, and these principles are applied in a wide range of careers.

In this text, you will begin to explore the history of the formal study of physics, beginning with natural philosophy and the ancient Greeks, and leading up through a review of Sir Isaac Newton and the laws of physics that bear his name. You will also be introduced to the standards scientists use when they study physical quantities and the interrelated system of measurements most of the scientific community uses to communicate in a single mathematical language. Finally, you will study the limits of our ability to be accurate and precise, and the reasons scientists go to painstaking lengths to be as clear as possible regarding their own limitations.

Physics: An Introduction

- Explain the difference between a principle and a law.
- Explain the difference between a model and a theory.

The flight formations of migratory birds such as Canada geese are governed by the laws of physics. (credit: David Merrett)

The physical universe is enormously complex in its detail. Every day, each of us observes a great variety of objects and phenomena. Over the centuries, the curiosity of the human race has led us collectively to explore and catalog a tremendous wealth of information. From the flight of birds to the colors of flowers, from lightning to gravity, from quarks to clusters of galaxies, from the flow of time to the mystery of the creation of the universe, we have asked questions and assembled huge arrays of facts. In the face of all these details, we have discovered that a surprisingly small and unified set of physical laws can explain what we observe. As humans, we make generalizations and seek order. We have found that nature is remarkably cooperative—it exhibits the *underlying order and simplicity* we so value.

It is the underlying order of nature that makes science in general, and physics in particular, so enjoyable to study. For example, what do a bag of chips and a car battery have in common? Both contain energy that can be

converted to other forms. The law of conservation of energy (which says that energy can change form but is never lost) ties together such topics as food calories, batteries, heat, light, and watch springs. Understanding this law makes it easier to learn about the various forms energy takes and how they relate to one another. Apparently unrelated topics are connected through broadly applicable physical laws, permitting an understanding beyond just the memorization of lists of facts.

The unifying aspect of physical laws and the basic simplicity of nature form the underlying themes of this text. In learning to apply these laws, you will, of course, study the most important topics in physics. More importantly, you will gain analytical abilities that will enable you to apply these laws far beyond the scope of what can be included in a single book. These analytical skills will help you to excel academically, and they will also help you to think critically in any professional career you choose to pursue. This module discusses the realm of physics (to define what physics is), some applications of physics (to illustrate its relevance to other disciplines), and more precisely what constitutes a physical law (to illuminate the importance of experimentation to theory).

Science and the Realm of Physics

Science consists of the theories and laws that are the general truths of nature as well as the body of knowledge they encompass. Scientists are continually trying to expand this body of knowledge and to perfect the expression of the laws that describe it. **Physics** is concerned with describing the interactions of energy, matter, space, and time, and it is especially interested in what fundamental mechanisms underlie every phenomenon. The concern for describing the basic phenomena in nature essentially defines the *realm of physics*.

Physics aims to describe the function of everything around us, from the movement of tiny charged particles to the motion of people, cars, and spaceships. In fact, almost everything around you can be described quite accurately by the laws of physics. Consider a smart phone ([link]). Physics describes how electricity interacts with the various circuits inside the device. This knowledge helps engineers select the appropriate materials and

circuit layout when building the smart phone. Next, consider a GPS system. Physics describes the relationship between the speed of an object, the distance over which it travels, and the time it takes to travel that distance. When you use a GPS device in a vehicle, it utilizes these physics equations to determine the travel time from one location to another.

The Apple "iPhone" is a common smart phone with a GPS function. Physics describes the way that electricity flows through the circuits of this device. Engineers use their knowledge of physics to construct an

iPhone with features that consumers will enjoy. One specific feature of an iPhone is the GPS function. GPS uses physics equations to determine the driving time between two locations on a map. (credit: @gletham GIS, Social, Mobile Tech Images)

Applications of Physics

You need not be a scientist to use physics. On the contrary, knowledge of physics is useful in everyday situations as well as in nonscientific professions. It can help you understand how microwave ovens work, why metals should not be put into them, and why they might affect pacemakers. (See [link] and [link].) Physics allows you to understand the hazards of radiation and rationally evaluate these hazards more easily. Physics also explains the reason why a black car radiator helps remove heat in a car engine, and it explains why a white roof helps keep the inside of a house cool. Similarly, the operation of a car's ignition system as well as the transmission of electrical signals through our body's nervous system are

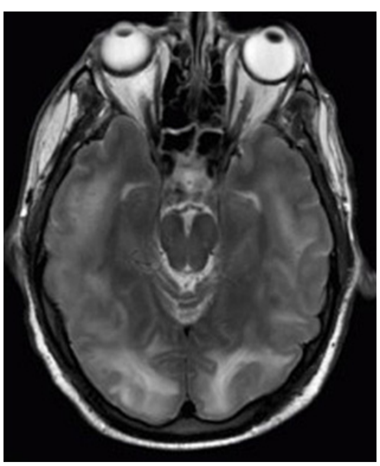
much easier to understand when you think about them in terms of basic physics.

Physics is the foundation of many important disciplines and contributes directly to others. Chemistry, for example—since it deals with the interactions of atoms and molecules—is rooted in atomic and molecular physics. Most branches of engineering are applied physics. In architecture, physics is at the heart of structural stability, and is involved in the acoustics, heating, lighting, and cooling of buildings. Parts of geology rely heavily on physics, such as radioactive dating of rocks, earthquake analysis, and heat transfer in the Earth. Some disciplines, such as biophysics and geophysics, are hybrids of physics and other disciplines.

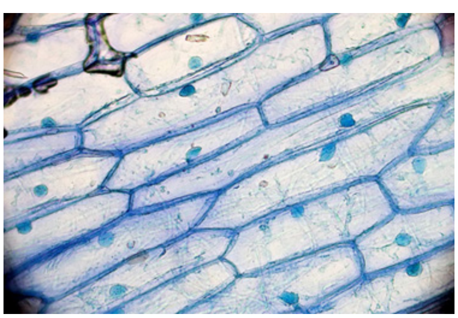
Physics has many applications in the biological sciences. On the microscopic level, it helps describe the properties of cell walls and cell membranes ([link] and [link]). On the macroscopic level, it can explain the heat, work, and power associated with the human body. Physics is involved in medical diagnostics, such as x-rays, magnetic resonance imaging (MRI), and ultrasonic blood flow measurements. Medical therapy sometimes directly involves physics; for example, cancer radiotherapy uses ionizing radiation. Physics can also explain sensory phenomena, such as how musical instruments make sound, how the eye detects color, and how lasers can transmit information.

It is not necessary to formally study all applications of physics. What is most useful is knowledge of the basic laws of physics and a skill in the analytical methods for applying them. The study of physics also can improve your problem-solving skills. Furthermore, physics has retained the most basic aspects of science, so it is used by all of the sciences, and the study of physics makes other sciences easier to understand.

The laws of physics help us understand how common appliances work. For example, the laws of physics can help explain how microwave ovens heat up food, and they also help us understand why it is dangerous to place metal objects in a microwave oven. (credit: MoneyBlogNewz)

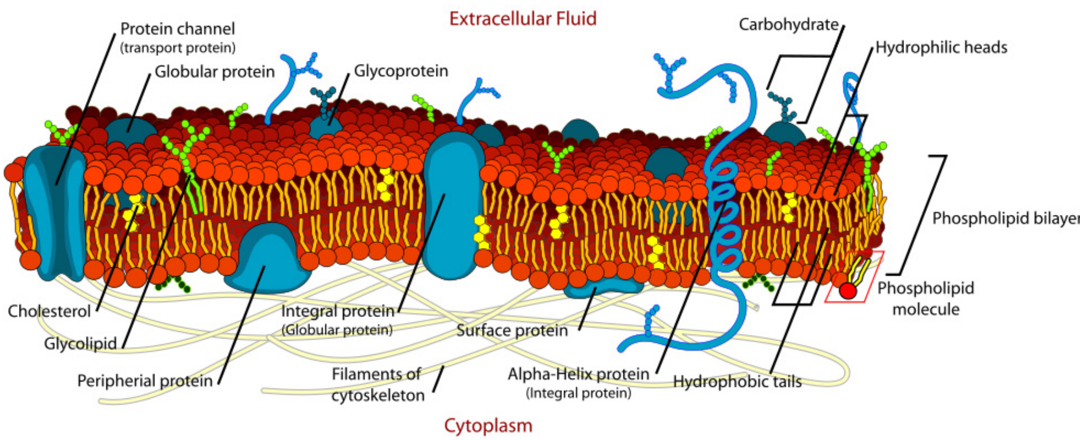


These two applications of physics have more in common than meets the eye. Microwave ovens use electromagnetic waves to heat food. Magnetic resonance imaging (MRI) also uses electromagnetic waves to yield an image of the brain, from which the exact location of tumors can be determined. (credit: Rashmi Chawla, Daniel Smith, and Paul E. Marik)



Physics, chemistry,

and biology help describe the properties of cell walls in plant cells, such as the onion cells seen here. (credit: Umberto Salvagnin)

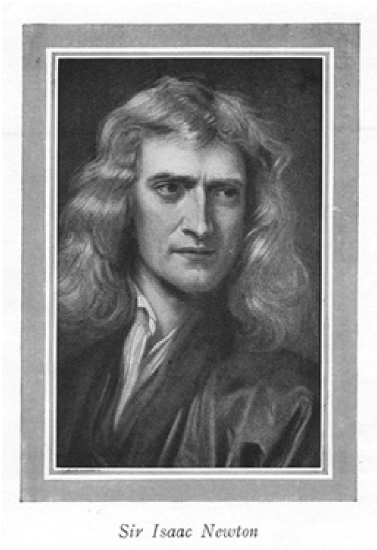


An artist’s rendition of the the structure of a cell membrane. Membranes form the boundaries of animal cells and are complex in structure and function. Many of the most fundamental properties of life, such as the firing of nerve cells, are related to membranes. The disciplines of biology, chemistry, and physics all help us understand the membranes of animal cells. (credit: Mariana Ruiz)

Models, Theories, and Laws; The Role of Experimentation

The laws of nature are concise descriptions of the universe around us; they are human statements of the underlying laws or rules that all natural processes follow. Such laws are intrinsic to the universe; humans did not

create them and so cannot change them. We can only discover and understand them. Their discovery is a very human endeavor, with all the elements of mystery, imagination, struggle, triumph, and disappointment inherent in any creative effort. (See [link] and [link].) The cornerstone of discovering natural laws is observation; science must describe the universe as it is, not as we may imagine it to be.



Isaac Newton (1642–1727) was very reluctant to publish his revolutionary work and had to be convinced to do so. In his later years, he stepped down from his academic post and became exchequer of the Royal Mint. He took this post

seriously, inventing reeding (or creating ridges) on the edge of coins to prevent unscrupulous people from trimming the silver off of them before using them as currency. (credit: Arthur Shuster and Arthur E. Shipley: *Britain's Heritage of Science*. London, 1917.)

Marie Curie (1867–1934) sacrificed

monetary assets to help finance her early research and damaged her physical well-being with radiation exposure. She is the only person to win Nobel prizes in both physics and chemistry. One of her daughters also won a Nobel Prize. (credit: Wikimedia Commons)

We all are curious to some extent. We look around, make generalizations, and try to understand what we see—for example, we look up and wonder whether one type of cloud signals an oncoming storm. As we become serious about exploring nature, we become more organized and formal in collecting and analyzing data. We attempt greater precision, perform controlled experiments (if we can), and write down ideas about how the data may be organized and unified. We then formulate models, theories, and laws based on the data we have collected and analyzed to generalize and communicate the results of these experiments.

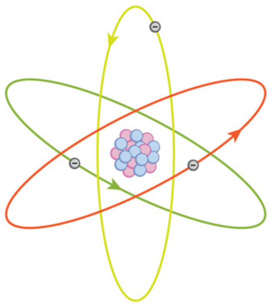
A **model** is a representation of something that is often too difficult (or impossible) to display directly. While a model is justified with experimental proof, it is only accurate under limited situations. An example is the planetary model of the atom in which electrons are pictured as orbiting the

nucleus, analogous to the way planets orbit the Sun. (See [link].) We cannot observe electron orbits directly, but the mental image helps explain the observations we can make, such as the emission of light from hot gases (atomic spectra). Physicists use models for a variety of purposes. For example, models can help physicists analyze a scenario and perform a calculation, or they can be used to represent a situation in the form of a computer simulation. A **theory** is an explanation for patterns in nature that is supported by scientific evidence and verified multiple times by various groups of researchers. Some theories include models to help visualize phenomena, whereas others do not. Newton’s theory of gravity, for example, does not require a model or mental image, because we can observe the objects directly with our own senses. The kinetic theory of gases, on the other hand, is a model in which a gas is viewed as being composed of atoms and molecules. Atoms and molecules are too small to be observed directly with our senses—thus, we picture them mentally to understand what our instruments tell us about the behavior of gases.

A **law** uses concise language to describe a generalized pattern in nature that is supported by scientific evidence and repeated experiments. Often, a law can be expressed in the form of a single mathematical equation. Laws and theories are similar in that they are both scientific statements that result from a tested hypothesis and are supported by scientific evidence. However, the designation *law* is reserved for a concise and very general statement that describes phenomena in nature, such as the law that energy is conserved during any process, or Newton’s second law of motion, which relates force, mass, and acceleration by the simple equation $\mathbf{F} = m\mathbf{a}$. A theory, in contrast, is a less concise statement of observed phenomena. For example, the Theory of Evolution and the Theory of Relativity cannot be expressed concisely enough to be considered a law. The biggest difference between a law and a theory is that a theory is much more complex and dynamic. A law describes a single action, whereas a theory explains an entire group of related phenomena. And, whereas a law is a postulate that forms the foundation of the scientific method, a theory is the end result of that process.

Less broadly applicable statements are usually called principles (such as Pascal’s principle, which is applicable only in fluids), but the distinction

between laws and principles often is not carefully made.



What is a model? This planetary model of the atom shows electrons orbiting the nucleus. It is a drawing that we use to form a mental image of the atom that we cannot see directly with our eyes because it is too small.

Note:
Models, Theories, and Laws
Models, theories, and laws are used to help scientists analyze the data they have already collected. However, often after a model, theory, or law has been developed, it points scientists toward new discoveries they would not otherwise have made.

The models, theories, and laws we devise sometimes *imply the existence of objects or phenomena as yet unobserved.* These predictions are remarkable triumphs and tributes to the power of science. It is the underlying order in the universe that enables scientists to make such spectacular predictions. However, if *experiment* does not verify our predictions, then the theory or law is wrong, no matter how elegant or convenient it is. Laws can never be known with absolute certainty because it is impossible to perform every imaginable experiment in order to confirm a law in every possible scenario. Physicists operate under the assumption that all scientific laws and theories are valid until a counterexample is observed. If a good-quality, verifiable experiment contradicts a well-established law, then the law must be modified or overthrown completely.

The study of science in general and physics in particular is an adventure much like the exploration of uncharted ocean. Discoveries are made; models, theories, and laws are formulated; and the beauty of the physical universe is made more sublime for the insights gained.

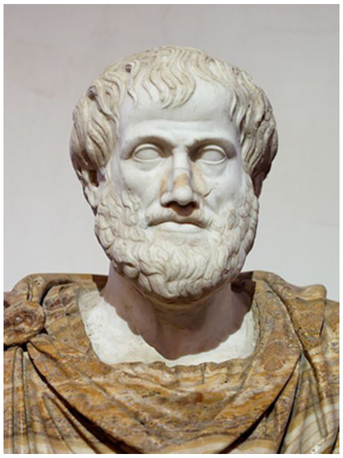
Note:
The Scientific Method
As scientists inquire and gather information about the world, they follow a process called the **scientific method**. This process typically begins with an observation and question that the scientist will research. Next, the scientist

typically performs some research about the topic and then devises a hypothesis. Then, the scientist will test the hypothesis by performing an experiment. Finally, the scientist analyzes the results of the experiment and draws a conclusion. Note that the scientific method can be applied to many situations that are not limited to science, and this method can be modified to suit the situation.

Consider an example. Let us say that you try to turn on your car, but it will not start. You undoubtedly wonder: Why will the car not start? You can follow a scientific method to answer this question. First off, you may perform some research to determine a variety of reasons why the car will not start. Next, you will state a hypothesis. For example, you may believe that the car is not starting because it has no engine oil. To test this, you open the hood of the car and examine the oil level. You observe that the oil is at an acceptable level, and you thus conclude that the oil level is not contributing to your car issue. To troubleshoot the issue further, you may devise a new hypothesis to test and then repeat the process again.

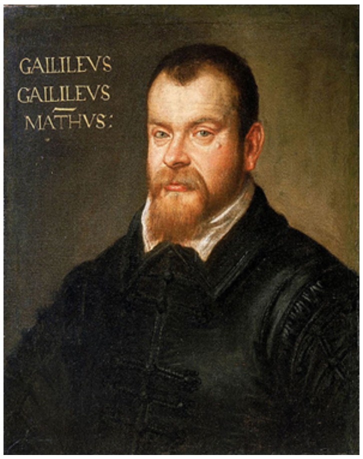
The Evolution of Natural Philosophy into Modern Physics

Physics was not always a separate and distinct discipline. It remains connected to other sciences to this day. The word *physics* comes from Greek, meaning nature. The study of nature came to be called “natural philosophy.” From ancient times through the Renaissance, natural philosophy encompassed many fields, including astronomy, biology, chemistry, physics, mathematics, and medicine. Over the last few centuries, the growth of knowledge has resulted in ever-increasing specialization and branching of natural philosophy into separate fields, with physics retaining the most basic facets. (See [link], [link], and [link].) Physics as it developed from the Renaissance to the end of the 19th century is called **classical physics**. It was transformed into modern physics by revolutionary discoveries made starting at the beginning of the 20th century.

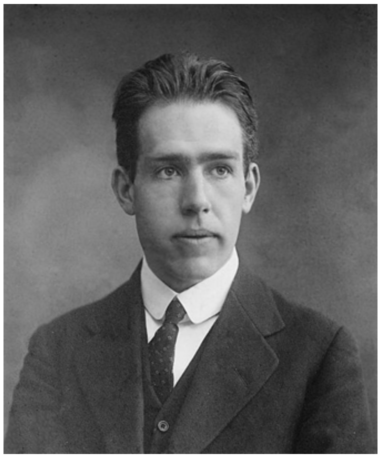


Over the centuries, natural philosophy has evolved into more specialized disciplines, as illustrated by the contributions of some of the greatest minds in history. The Greek philosopher **Aristotle** (384–322 B.C.) wrote on a broad range of topics including physics, animals, the soul, politics, and poetry. (credit: Jastrow

(2006)/Ludovisi Collection)



Galileo Galilei (1564–1642) laid the foundation of modern experimentation and made contributions in mathematics, physics, and astronomy. (credit: Domenico Tintoretto)

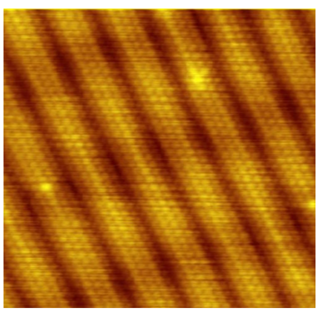


Niels Bohr (1885–1962) made fundamental contributions to the development of quantum mechanics, one part of modern physics. (credit: United States Library of Congress Prints and Photographs Division)

Classical physics is not an exact description of the universe, but it is an excellent approximation under the following conditions: Matter must be moving at speeds less than about 1% of the speed of light, the objects dealt with must be large enough to be seen with a microscope, and only weak gravitational fields, such as the field generated by the Earth, can be involved. Because humans live under such circumstances, classical physics seems intuitively reasonable, while many aspects of modern physics seem bizarre. This is why models are so useful in modern physics—they let us

conceptualize phenomena we do not ordinarily experience. We can relate to models in human terms and visualize what happens when objects move at high speeds or imagine what objects too small to observe with our senses might be like. For example, we can understand an atom's properties because we can picture it in our minds, although we have never seen an atom with our eyes. New tools, of course, allow us to better picture phenomena we cannot see. In fact, new instrumentation has allowed us in recent years to actually "picture" the atom.

Note:
Limits on the Laws of Classical Physics
For the laws of classical physics to apply, the following criteria must be met: Matter must be moving at speeds less than about 1% of the speed of light, the objects dealt with must be large enough to be seen with a microscope, and only weak gravitational fields (such as the field generated by the Earth) can be involved.



Using a scanning tunneling microscope (STM), scientists can see the individual atoms that

compose this sheet of gold. (credit: Erwinrossen)

Some of the most spectacular advances in science have been made in modern physics. Many of the laws of classical physics have been modified or rejected, and revolutionary changes in technology, society, and our view of the universe have resulted. Like science fiction, modern physics is filled with fascinating objects beyond our normal experiences, but it has the advantage over science fiction of being very real. Why, then, is the majority of this text devoted to topics of classical physics? There are two main reasons: Classical physics gives an extremely accurate description of the universe under a wide range of everyday circumstances, and knowledge of classical physics is necessary to understand modern physics.

Modern physics itself consists of the two revolutionary theories, relativity and quantum mechanics. These theories deal with the very fast and the very small, respectively. **Relativity** must be used whenever an object is traveling at greater than about 1% of the speed of light or experiences a strong gravitational field such as that near the Sun. **Quantum mechanics** must be used for objects smaller than can be seen with a microscope. The combination of these two theories is *relativistic quantum mechanics*, and it describes the behavior of small objects traveling at high speeds or experiencing a strong gravitational field. Relativistic quantum mechanics is the best universally applicable theory we have. Because of its mathematical complexity, it is used only when necessary, and the other theories are used whenever they will produce sufficiently accurate results. We will find, however, that we can do a great deal of modern physics with the algebra and trigonometry used in this text.

Exercise:

Check Your Understanding

Problem:

A friend tells you he has learned about a new law of nature. What can you know about the information even before your friend describes the law? How would the information be different if your friend told you he had learned about a scientific theory rather than a law?

Solution:

Without knowing the details of the law, you can still infer that the information your friend has learned conforms to the requirements of all laws of nature: it will be a concise description of the universe around us; a statement of the underlying rules that all natural processes follow. If the information had been a theory, you would be able to infer that the information will be a large-scale, broadly applicable generalization.

Note:
PhET Explorations: Equation Grapher
Learn about graphing polynomials. The shape of the curve changes as the constants are adjusted. View the curves for the individual terms (e.g. $y = bx$) to see how they add to generate the polynomial curve.
https://phet.colorado.edu/sims/equation-grapher/equation-grapher_en.html

Summary

- Science seeks to discover and describe the underlying order and simplicity in nature.
- Physics is the most basic of the sciences, concerning itself with energy, matter, space and time, and their interactions.
- Scientific laws and theories express the general truths of nature and the body of knowledge they encompass. These laws of nature are rules that all natural processes appear to follow.

Conceptual Questions

Exercise:

Problem:

Models are particularly useful in relativity and quantum mechanics, where conditions are outside those normally encountered by humans. What is a model?

Exercise:

Problem: How does a model differ from a theory?

Exercise:

Problem:

If two different theories describe experimental observations equally well, can one be said to be more valid than the other (assuming both use accepted rules of logic)?

Exercise:

Problem: What determines the validity of a theory?

Exercise:

Problem:

Certain criteria must be satisfied if a measurement or observation is to be believed. Will the criteria necessarily be as strict for an expected result as for an unexpected result?

Exercise:

Problem:

Can the validity of a model be limited, or must it be universally valid? How does this compare to the required validity of a theory or a law?

Exercise:

Problem:

Classical physics is a good approximation to modern physics under certain circumstances. What are they?

Exercise:

Problem: When is it *necessary* to use relativistic quantum mechanics?

Exercise:

Problem:

Can classical physics be used to accurately describe a satellite moving at a speed of 7500 m/s? Explain why or why not.

Glossary

classical physics
: physics that was developed from the Renaissance to the end of the 19th century

physics
: the science concerned with describing the interactions of energy, matter, space, and time; it is especially interested in what fundamental mechanisms underlie every phenomenon

model
: representation of something that is often too difficult (or impossible) to display directly

theory
: an explanation for patterns in nature that is supported by scientific evidence and verified multiple times by various groups of researchers

law
: a description, using concise language or a mathematical formula, a generalized pattern in nature that is supported by scientific evidence

and repeated experiments

scientific method

a method that typically begins with an observation and question that the scientist will research; next, the scientist typically performs some research about the topic and then devises a hypothesis; then, the scientist will test the hypothesis by performing an experiment; finally, the scientist analyzes the results of the experiment and draws a conclusion

modern physics

the study of relativity, quantum mechanics, or both

relativity

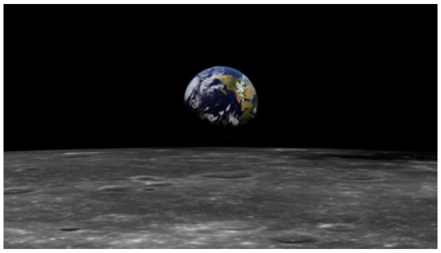
the study of objects moving at speeds greater than about 1% of the speed of light, or of objects being affected by a strong gravitational field

quantum mechanics

the study of objects smaller than can be seen with a microscope

Physical Quantities and Units

- Perform unit conversions both in the SI and English units.
- Explain the most common prefixes in the SI units and be able to write them in scientific notation.

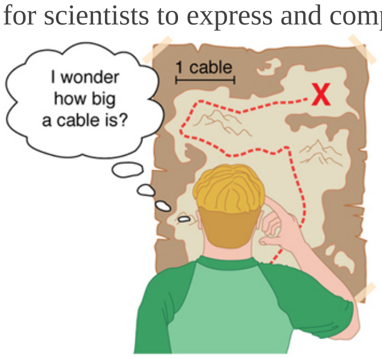


The distance from Earth to the Moon may seem immense, but it is just a tiny fraction of the distances from Earth to other celestial bodies. (credit: NASA)

The range of objects and phenomena studied in physics is immense. From the incredibly short lifetime of a nucleus to the age of the Earth, from the tiny sizes of sub-nuclear particles to the vast distance to the edges of the known universe, from the force exerted by a jumping flea to the force between Earth and the Sun, there are enough factors of 10 to challenge the imagination of even the most experienced scientist. Giving numerical values for physical quantities and equations for physical principles allows us to understand nature much more deeply than does qualitative description alone. To comprehend these vast ranges, we must also have accepted units in which to express them. And we shall find that (even in the potentially mundane discussion of meters, kilograms, and seconds) a profound simplicity of nature appears—all physical quantities can be expressed as combinations of only four fundamental physical quantities: length, mass, time, and electric current.

We define a **physical quantity** either by *specifying how it is measured* or by *stating how it is calculated* from other measurements. For example, we define distance and time by specifying methods for measuring them, whereas we define *average speed* by stating that it is calculated as distance traveled divided by time of travel.

Measurements of physical quantities are expressed in terms of **units**, which are standardized values. For example, the length of a race, which is a physical quantity, can be expressed in units of meters (for sprinters) or kilometers (for distance runners). Without standardized units, it would be extremely difficult for scientists to express and compare measured values in a meaningful way. (See [link].)



Distances given in unknown units are maddeningly useless.

There are two major systems of units used in the world: **SI units** (also known as the metric system) and **English units** (also known as the customary or imperial system). **English units** were historically used in nations once ruled by the British Empire and are still widely used in the United States. Virtually every other country in the world now uses SI units as the standard; the metric system is also the standard system agreed upon by scientists and mathematicians. The acronym “SI” is derived from the French *Système International*.

SI Units: Fundamental and Derived Units

[link] gives the fundamental SI units that are used throughout this textbook. This text uses non-SI units in a few applications where they are in very common use, such as the measurement of blood pressure in millimeters of mercury (mm Hg). Whenever non-SI units are discussed, they will be tied to SI units through conversions.

Length	Mass	Time	Electric Current
meter (m)	kilogram (kg)	second (s)	ampere (A)

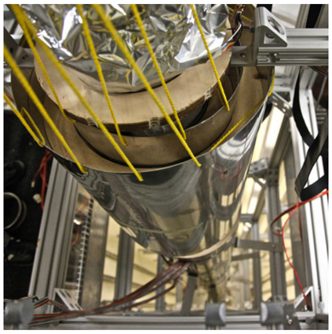
Fundamental SI Units

It is an intriguing fact that some physical quantities are more fundamental than others and that the most fundamental physical quantities can be defined *only* in terms of the procedure used to measure them. The units in which they are measured are thus called **fundamental units**. In this textbook, the fundamental physical quantities are taken to be length, mass, time, and electric current. (Note that electric current will not be introduced until much later in this text.) All other physical quantities, such as force and electric charge, can be expressed as algebraic combinations of length, mass, time, and current (for example, speed is length divided by time); these units are called **derived units**.

Units of Time, Length, and Mass: The Second, Meter, and Kilogram

The Second

The SI unit for time, the **second**(abbreviated s), has a long history. For many years it was defined as 1/86,400 of a mean solar day. More recently, a new standard was adopted to gain greater accuracy and to define the second in terms of a non-varying, or constant, physical phenomenon (because the solar day is getting longer due to very gradual slowing of the Earth’s rotation). Cesium atoms can be made to vibrate in a very steady way, and these vibrations can be readily observed and counted. In 1967 the second was redefined as the time required for 9,192,631,770 of these vibrations. (See [link].) Accuracy in the fundamental units is essential, because all measurements are ultimately expressed in terms of fundamental units and can be no more accurate than are the fundamental units themselves.



An atomic clock such as this one uses the vibrations of cesium atoms to keep time to a precision of better than a microsecond per year. The fundamental unit of time, the second, is based on such clocks. This image is looking down from the top of an atomic fountain nearly 30 feet tall! (credit: Steve Jurvetson/Flickr)

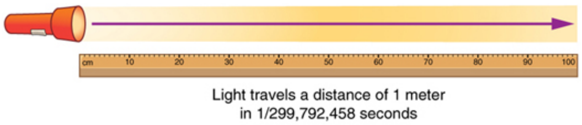
The Meter

The SI unit for length is the **meter** (abbreviated m); its definition has also changed over time to become more accurate and precise. The meter was first defined in 1791 as 1/10,000,000 of the distance from the equator to the North Pole. This measurement was improved in 1889 by redefining the meter to be the distance between two engraved lines on a platinum-iridium bar now kept near Paris. By 1960, it had become possible to define the meter even more accurately in terms of the wavelength of light, so it was again redefined as 1,650,763.73 wavelengths of orange light emitted by krypton atoms. In 1983, the meter was given its present definition (partly for greater accuracy) as the distance light travels in a vacuum in 1/299,792,458 of a second. (See [link].) This change defines the speed of light to be exactly 299,792,458 meters per second. The length of the meter will change if the speed of light is someday measured with greater accuracy.

The Kilogram

The SI unit for mass is the **kilogram** (abbreviated kg); it is defined to be the mass of a platinum-iridium cylinder kept with the old meter standard at the International Bureau of Weights and Measures near Paris. Exact replicas of the standard kilogram are also kept at the United States' National Institute of Standards

and Technology, or NIST, located in Gaithersburg, Maryland outside of Washington D.C., and at other locations around the world. The determination of all other masses can be ultimately traced to a comparison with the standard mass.



The meter is defined to be the distance light travels in 1/299,792,458 of a second in a vacuum. Distance traveled is speed multiplied by time.

Electric current and its accompanying unit, the ampere, will be introduced in Introduction to Electric Current, Resistance, and Ohm's Law when electricity and magnetism are covered. The initial modules in this textbook are concerned with mechanics, fluids, heat, and waves. In these subjects all pertinent physical quantities can be expressed in terms of the fundamental units of length, mass, and time.

Metric Prefixes

SI units are part of the **metric system**. The metric system is convenient for scientific and engineering calculations because the units are categorized by factors of 10. [link] gives metric prefixes and symbols used to denote various factors of 10.

Metric systems have the advantage that conversions of units involve only powers of 10. There are 100 centimeters in a meter, 1000 meters in a kilometer, and so on. In nonmetric systems, such as the system of U.S. customary units, the relationships are not as simple—there are 12 inches in a foot, 5280 feet in a mile, and so on. Another advantage of the metric system is that the same unit can be used over extremely large ranges of values simply by using an appropriate metric prefix. For example, distances in meters are suitable in construction, while distances in kilometers are appropriate for air travel, and the tiny measure of nanometers are convenient in optical design. With the metric system there is no need to invent new units for particular applications.

The term **order of magnitude** refers to the scale of a value expressed in the metric system. Each power of 10 in the metric system represents a different order of magnitude. For example, 10^1, 10^2, 10^3, and so forth are all different orders of magnitude. All quantities that can be expressed as a product of a specific power of 10 are said to be of the *same* order of magnitude. For example, the number 800 can be written as 8×10^2, and the number 450 can be written as 4.5×10^2. Thus, the numbers 800 and 450 are of the same order of magnitude: 10^2. Order of magnitude can be thought of as a ballpark estimate for the scale of a value. The diameter of an atom is on the order of $10^{-9}\ \text{m}$, while the diameter of the Sun is on the order of $10^9\ \text{m}$.

Note:
The Quest for Microscopic Standards for Basic Units

The fundamental units described in this chapter are those that produce the greatest accuracy and precision in measurement. There is a sense among physicists that, because there is an underlying microscopic substructure to matter, it would be most satisfying to base our standards of measurement on microscopic objects and fundamental physical phenomena such as the speed of light. A microscopic standard has been accomplished for the standard of time, which is based on the oscillations of the cesium atom.

The standard for length was once based on the wavelength of light (a small-scale length) emitted by a certain type of atom, but it has been supplanted by the more precise measurement of the speed of light. If it becomes possible to measure the mass of atoms or a particular arrangement of atoms such as a silicon sphere to greater precision than the kilogram standard, it may become possible to base mass measurements on the small scale. There are also possibilities that electrical phenomena on the small scale may someday allow us to base a unit of charge on the charge of electrons and protons, but at present current and charge are related to large-scale currents and forces between wires.

Prefix	**Symbol**	**Value[footnote] See Appendix A for a discussion of powers of 10.**	**Example (some are approximate)**			
exa	E	10^{18}	exameter	Em	10^{18} m	distance light travels in a century
peta	P	10^{15}	petasecond	Ps	10^{15} s	30 million years
tera	T	10^{12}	terawatt	TW	10^{12} W	powerful laser output
giga	G	10^{9}	gigahertz	GHz	10^{9} Hz	a microwave frequency
mega	M	10^{6}	megacurie	MCi	10^{6} Ci	high radioactivity
kilo	k	10^{3}	kilometer	km	10^{3} m	about 6/10 mile
hecto	h	10^{2}	hectoliter	hL	10^{2} L	26 gallons

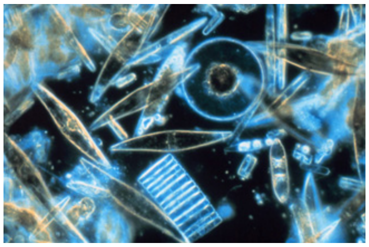
Prefix	Symbol	Value[footnote] See Appendix A for a discussion of powers of 10.	Example (some are approximate)			
deka	da	10^1	dekagram	dag	10^1 g	teaspoon of butter
—	—	10^0 (=1)				
deci	d	10^{-1}	deciliter	dL	10^{-1} L	less than half a soda
centi	c	10^{-2}	centimeter	cm	10^{-2} m	fingertip thickness
milli	m	10^{-3}	millimeter	mm	10^{-3} m	flea at its shoulders
micro	µ	10^{-6}	micrometer	µm	10^{-6} m	detail in microscope
nano	n	10^{-9}	nanogram	ng	10^{-9} g	small speck of dust
pico	p	10^{-12}	picofarad	pF	10^{-12} F	small capacitor in radio
femto	f	10^{-15}	femtometer	fm	10^{-15} m	size of a proton
atto	a	10^{-18}	attosecond	as	10^{-18} s	time light crosses an atom

Metric Prefixes for Powers of 10 and their Symbols

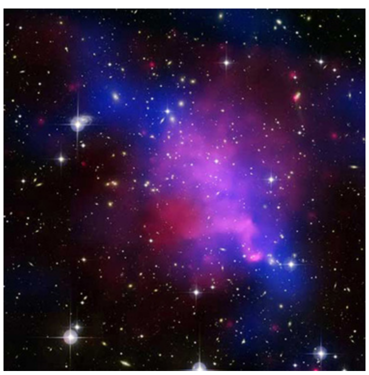
Known Ranges of Length, Mass, and Time

The vastness of the universe and the breadth over which physics applies are illustrated by the wide range of examples of known lengths, masses, and times in [link]. Examination of this table will give you some

feeling for the range of possible topics and numerical values. (See [link] and [link].)



Tiny phytoplankton swims among crystals of ice in the Antarctic Sea. They range from a few micrometers to as much as 2 millimeters in length. (credit: Prof. Gordon T. Taylor, Stony Brook University; NOAA Corps Collections)



Galaxies collide 2.4 billion light years away from Earth. The tremendous range of observable phenomena in nature challenges the imagination. (credit: NASA/CXC/UVic./A. Mahdavi et al. Optical/lensing: CFHT/UVic./H. Hoekstra et al.)

Unit Conversion and Dimensional Analysis

It is often necessary to convert from one type of unit to another. For example, if you are reading a European cookbook, some quantities may be expressed in units of liters and you need to convert them to cups. Or, perhaps you are reading walking directions from one location to another and you are interested in how many miles you will be walking. In this case, you will need to convert units of feet to miles.

Let us consider a simple example of how to convert units. Let us say that we want to convert 80 meters (m) to kilometers (km).

The first thing to do is to list the units that you have and the units that you want to convert to. In this case, we have units in *meters* and we want to convert to *kilometers*.

Next, we need to determine a **conversion factor** relating meters to kilometers. A conversion factor is a ratio expressing how many of one unit are equal to another unit. For example, there are 12 inches in 1 foot, 100 centimeters in 1 meter, 60 seconds in 1 minute, and so on. In this case, we know that there are 1,000 meters in 1 kilometer.

Now we can set up our unit conversion. We will write the units that we have and then multiply them by the conversion factor so that the units cancel out, as shown:
Equation:

$$80\,\cancel{\text{m}} \times \frac{1\text{ km}}{1000\,\cancel{\text{m}}} = 0.080\text{ km}.$$

Note that the unwanted m unit cancels, leaving only the desired km unit. You can use this method to convert between any types of unit.

Click [link] for a more complete list of conversion factors.

Lengths in meters		Masses in kilograms (more precise values in parentheses)		Times in seconds (more precise values in parentheses)	
10^{-18}	Present experimental limit to smallest observable detail	10^{-30}	Mass of an electron $\left(9.11 \times 10^{-31}\text{ kg}\right)$	10^{-23}	Time for light to cross a proton
10^{-15}	Diameter of a proton	10^{-27}	Mass of a hydrogen atom $\left(1.67 \times 10^{-27}\text{ kg}\right)$	10^{-22}	Mean life of an extremely unstable nucleus

Lengths in meters		Masses in kilograms (more precise values in parentheses)		Times in seconds (more precise values in parentheses)	
10^{-14}	Diameter of a uranium nucleus	10^{-15}	Mass of a bacterium	10^{-15}	Time for one oscillation of visible light
10^{-10}	Diameter of a hydrogen atom	10^{-5}	Mass of a mosquito	10^{-13}	Time for one vibration of an atom in a solid
10^{-8}	Thickness of membranes in cells of living organisms	10^{-2}	Mass of a hummingbird	10^{-8}	Time for one oscillation of an FM radio wave
10^{-6}	Wavelength of visible light	1	Mass of a liter of water (about a quart)	10^{-3}	Duration of a nerve impulse
10^{-3}	Size of a grain of sand	10^{2}	Mass of a person	1	Time for one heartbeat
1	Height of a 4-year-old child	10^{3}	Mass of a car	10^{5}	One day $(8.64 \times 10^{4}\ \text{s})$
10^{2}	Length of a football field	10^{8}	Mass of a large ship	10^{7}	One year (y) $(3.16 \times 10^{7}\ \text{s})$
10^{4}	Greatest ocean depth	10^{12}	Mass of a large iceberg	10^{9}	About half the life expectancy of a human
10^{7}	Diameter of the Earth	10^{15}	Mass of the nucleus of a comet	10^{11}	Recorded history
10^{11}	Distance from the Earth to the Sun	10^{23}	Mass of the Moon $(7.35 \times 10^{22}\ \text{kg})$	10^{17}	Age of the Earth
10^{16}	Distance traveled by light in 1 year (a light year)	10^{25}	Mass of the Earth $(5.97 \times 10^{24}\ \text{kg})$	10^{18}	Age of the universe
10^{21}	Diameter of the Milky Way galaxy	10^{30}	Mass of the Sun $(1.99 \times 10^{30}\ \text{kg})$		

Lengths in meters		Masses in kilograms (more precise values in parentheses)		Times in seconds (more precise values in parentheses)	
10^{22}	Distance from the Earth to the nearest large galaxy (Andromeda)	10^{42}	Mass of the Milky Way galaxy (current upper limit)		
10^{26}	Distance from the Earth to the edges of the known universe	10^{53}	Mass of the known universe (current upper limit)		

Approximate Values of Length, Mass, and Time

Example:
Unit Conversions: A Short Drive Home
Suppose that you drive the 10.0 km from your university to home in 20.0 min. Calculate your average speed (a) in kilometers per hour (km/h) and (b) in meters per second (m/s). (Note: Average speed is distance traveled divided by time of travel.)
Strategy
First we calculate the average speed using the given units. Then we can get the average speed into the desired units by picking the correct conversion factor and multiplying by it. The correct conversion factor is the one that cancels the unwanted unit and leaves the desired unit in its place.
Solution for (a)
(1) Calculate average speed. Average speed is distance traveled divided by time of travel. (Take this definition as a given for now—average speed and other motion concepts will be covered in a later module.) In equation form,
Equation:

$$\text{average speed} = \frac{\text{distance}}{\text{time}}.$$

(2) Substitute the given values for distance and time.
Equation:

$$\text{average speed} = \frac{10.0 \text{ km}}{20.0 \text{ min}} = 0.500 \frac{\text{km}}{\text{min}}.$$

(3) Convert km/min to km/h: multiply by the conversion factor that will cancel minutes and leave hours. That conversion factor is 60 min/hr. Thus,
Equation:

$$\text{average speed} = 0.500 \frac{\text{km}}{\text{min}} \times \frac{60 \text{ min}}{1 \text{ h}} = 30.0 \frac{\text{km}}{\text{h}}.$$

Discussion for (a)
To check your answer, consider the following:

(1) Be sure that you have properly cancelled the units in the unit conversion. If you have written the unit conversion factor upside down, the units will not cancel properly in the equation. If you accidentally get the ratio upside down, then the units will not cancel; rather, they will give you the wrong units as follows:

Equation:

$$\frac{\text{km}}{\text{min}} \times \frac{1\text{ hr}}{60\text{ min}} = \frac{1}{60}\frac{\text{km}\cdot\text{hr}}{\text{min}^2},$$

which are obviously not the desired units of km/h.

(2) Check that the units of the final answer are the desired units. The problem asked us to solve for average speed in units of km/h and we have indeed obtained these units.

(3) Check the significant figures. Because each of the values given in the problem has three significant figures, the answer should also have three significant figures. The answer 30.0 km/hr does indeed have three significant figures, so this is appropriate. Note that the significant figures in the conversion factor are not relevant because an hour is *defined* to be 60 minutes, so the precision of the conversion factor is perfect.

(4) Next, check whether the answer is reasonable. Let us consider some information from the problem—if you travel 10 km in a third of an hour (20 min), you would travel three times that far in an hour. The answer does seem reasonable.

Solution for (b)

There are several ways to convert the average speed into meters per second.

(1) Start with the answer to (a) and convert km/h to m/s. Two conversion factors are needed—one to convert hours to seconds, and another to convert kilometers to meters.

(2) Multiplying by these yields

Equation:

$$\text{Average speed} = 30.0\frac{\text{km}}{\text{h}} \times \frac{1\text{ h}}{3{,}600\text{ s}} \times \frac{1{,}000\text{ m}}{1\text{ km}},$$

Equation:

$$\text{Average speed} = 8.33\frac{\text{m}}{\text{s}}.$$

Discussion for (b)

If we had started with 0.500 km/min, we would have needed different conversion factors, but the answer would have been the same: 8.33 m/s.

You may have noted that the answers in the worked example just covered were given to three digits. Why? When do you need to be concerned about the number of digits in something you calculate? Why not write down all the digits your calculator produces? The module Accuracy, Precision, and Significant Figures will help you answer these questions.

Note:

Nonstandard Units

While there are numerous types of units that we are all familiar with, there are others that are much more obscure. For example, a **firkin** is a unit of volume that was once used to measure beer. One firkin equals about 34 liters. To learn more about nonstandard units, use a dictionary or encyclopedia to research different "weights and measures." Take note of any unusual units, such as a barleycorn, that are not listed in the text. Think about how the unit is defined and state its relationship to SI units.

Exercise:
Check Your Understanding

Problem:

Some hummingbirds beat their wings more than 50 times per second. A scientist is measuring the time it takes for a hummingbird to beat its wings once. Which fundamental unit should the scientist use to describe the measurement? Which factor of 10 is the scientist likely to use to describe the motion precisely? Identify the metric prefix that corresponds to this factor of 10.

Solution:

The scientist will measure the time between each movement using the fundamental unit of seconds. Because the wings beat so fast, the scientist will probably need to measure in milliseconds, or 10^{-3} seconds. (50 beats per second corresponds to 20 milliseconds per beat.)

Exercise:
Check Your Understanding

Problem:

One cubic centimeter is equal to one milliliter. What does this tell you about the different units in the SI metric system?

Solution:

The fundamental unit of length (meter) is probably used to create the derived unit of volume (liter). The measure of a milliliter is dependent on the measure of a centimeter.

Summary

- Physical quantities are a characteristic or property of an object that can be measured or calculated from other measurements.
- Units are standards for expressing and comparing the measurement of physical quantities. All units can be expressed as combinations of four fundamental units.
- The four fundamental units we will use in this text are the meter (for length), the kilogram (for mass), the second (for time), and the ampere (for electric current). These units are part of the metric system, which uses powers of 10 to relate quantities over the vast ranges encountered in nature.
- The four fundamental units are abbreviated as follows: meter, m; kilogram, kg; second, s; and ampere, A. The metric system also uses a standard set of prefixes to denote each order of magnitude greater than or lesser than the fundamental unit itself.
- Unit conversions involve changing a value expressed in one type of unit to another type of unit. This is done by using conversion factors, which are ratios relating equal quantities of different units.

Conceptual Questions

Exercise:

Problem: Identify some advantages of metric units.

Problems & Exercises

Exercise:

Problem:

The speed limit on some interstate highways is roughly 100 km/h. (a) What is this in meters per second? (b) How many miles per hour is this?

Solution:

a. 27.8 m/s
b. 62.1 mph

Exercise:

Problem:

A car is traveling at a speed of 33 m/s. (a) What is its speed in kilometers per hour? (b) Is it exceeding the 90 km/h speed limit?

Exercise:

Problem:

Show that $1.0 \text{ m/s} = 3.6 \text{ km/h}$. Hint: Show the explicit steps involved in converting $1.0 \text{ m/s} = 3.6 \text{ km/h}.$

Solution:

$$\frac{1.0 \text{ m}}{\text{s}} = \frac{1.0 \text{ m}}{\text{s}} \times \frac{3600 \text{ s}}{1 \text{ hr}} \times \frac{1 \text{ km}}{1000 \text{ m}}$$

$$= 3.6 \text{ km/h}.$$

Exercise:

Problem:

American football is played on a 100-yd-long field, excluding the end zones. How long is the field in meters? (Assume that 1 meter equals 3.281 feet.)

Exercise:

Problem:

Soccer fields vary in size. A large soccer field is 115 m long and 85 m wide. What are its dimensions in feet and inches? (Assume that 1 meter equals 3.281 feet.)

Solution:

length: $377 \text{ ft}; 4.53 \times 10^3 \text{ in}.$ width: $280 \text{ ft}; 3.3 \times 10^3 \text{ in}.$

Exercise:

Problem:

What is the height in meters of a person who is 6 ft 1.0 in. tall? (Assume that 1 meter equals 39.37 in.)

Exercise:

Problem:

Mount Everest, at 29,028 feet, is the tallest mountain on the Earth. What is its height in kilometers? (Assume that 1 kilometer equals 3,281 feet.)

Solution:

8.847 km

Exercise:

Problem: The speed of sound is measured to be 342 m/s on a certain day. What is this in km/h?

Exercise:

Problem:

Tectonic plates are large segments of the Earth's crust that move slowly. Suppose that one such plate has an average speed of 4.0 cm/year. (a) What distance does it move in 1 s at this speed? (b) What is its speed in kilometers per million years?

Solution:

(a) $1.3 \times 10^{-9}\text{ m}$

(b) 40 km/My

Exercise:

Problem:

(a) Refer to [link] to determine the average distance between the Earth and the Sun. Then calculate the average speed of the Earth in its orbit in kilometers per second. (b) What is this in meters per second?

Glossary

physical quantity
: a characteristic or property of an object that can be measured or calculated from other measurements

units
: a standard used for expressing and comparing measurements

SI units
: the international system of units that scientists in most countries have agreed to use; includes units such as meters, liters, and grams

English units
: system of measurement used in the United States; includes units of measurement such as feet, gallons, and pounds

fundamental units
: units that can only be expressed relative to the procedure used to measure them

derived units
: units that can be calculated using algebraic combinations of the fundamental units

second
: the SI unit for time, abbreviated (s)

meter
: the SI unit for length, abbreviated (m)

kilogram
: the SI unit for mass, abbreviated (kg)

metric system
: a system in which values can be calculated in factors of 10

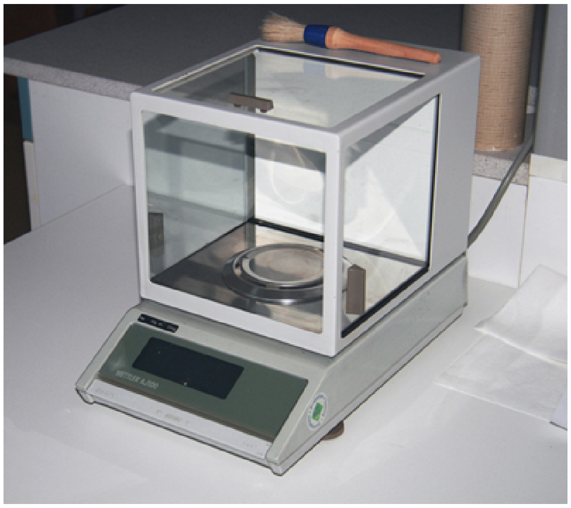
order of magnitude
: refers to the size of a quantity as it relates to a power of 10

conversion factor
: a ratio expressing how many of one unit are equal to another unit

Accuracy, Precision, and Significant Figures

- Determine the appropriate number of significant figures in both addition and subtraction, as well as multiplication and division calculations.
- Calculate the percent uncertainty of a measurement.

A double-pan mechanical balance is used to compare different masses. Usually an object with unknown mass is placed in one pan and objects of known mass are placed in the other pan. When the bar that connects the two pans is horizontal, then the masses in both pans are equal. The “known masses” are typically metal cylinders of standard mass such as 1 gram, 10 grams, and 100 grams. (credit: Serge Melki)



Many mechanical balances, such as double-pan balances, have been replaced by digital scales, which can typically measure the mass of an object more precisely. Whereas a mechanical balance may only read the mass of an object to the nearest tenth of a gram, many digital scales can measure the mass of an object up to the nearest thousandth of a gram. (credit: Karel Jakubec)

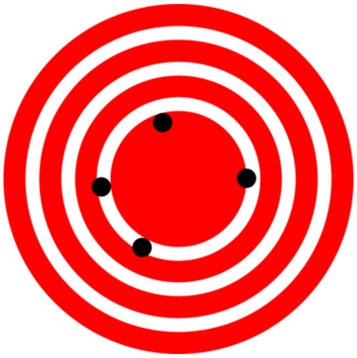
Accuracy and Precision of a Measurement

Science is based on observation and experiment—that is, on measurements. **Accuracy** is how close a measurement is to the correct value for that measurement. For example, let us say that you are measuring the length of standard computer paper. The packaging in which you purchased the paper states that it is 11.0 inches long. You measure the length of the paper three times and obtain the following measurements: 11.1 in., 11.2 in., and 10.9 in.

These measurements are quite accurate because they are very close to the correct value of 11.0 inches. In contrast, if you had obtained a measurement of 12 inches, your measurement would not be very accurate.

The **precision** of a measurement system is refers to how close the agreement is between repeated measurements (which are repeated under the same conditions). Consider the example of the paper measurements. The precision of the measurements refers to the spread of the measured values. One way to analyze the precision of the measurements would be to determine the range, or difference, between the lowest and the highest measured values. In that case, the lowest value was 10.9 in. and the highest value was 11.2 in. Thus, the measured values deviated from each other by at most 0.3 in. These measurements were relatively precise because they did not vary too much in value. However, if the measured values had been 10.9, 11.1, and 11.9, then the measurements would not be very precise because there would be significant variation from one measurement to another.

The measurements in the paper example are both accurate and precise, but in some cases, measurements are accurate but not precise, or they are precise but not accurate. Let us consider an example of a GPS system that is attempting to locate the position of a restaurant in a city. Think of the restaurant location as existing at the center of a bull's-eye target, and think of each GPS attempt to locate the restaurant as a black dot. In [link], you can see that the GPS measurements are spread out far apart from each other, but they are all relatively close to the actual location of the restaurant at the center of the target. This indicates a low precision, high accuracy measuring system. However, in [link], the GPS measurements are concentrated quite closely to one another, but they are far away from the target location. This indicates a high precision, low accuracy measuring system.



A GPS system attempts to locate a restaurant at the center of the bull’s-eye. The black dots represent each attempt to pinpoint the location of the restaurant. The dots are spread out quite far apart from one another, indicating low precision, but they are each rather close to the actual location of the restaurant, indicating high accuracy. (credit: Dark Evil)

In this figure, the dots are concentrated rather closely to one another, indicating high precision, but they are rather far away from the actual location of the restaurant, indicating low accuracy. (credit: Dark Evil)

Accuracy, Precision, and Uncertainty

The degree of accuracy and precision of a measuring system are related to the **uncertainty** in the measurements. Uncertainty is a quantitative measure of how much your measured values deviate from a standard or expected value. If your measurements are not very accurate or precise, then the

uncertainty of your values will be very high. In more general terms, uncertainty can be thought of as a disclaimer for your measured values. For example, if someone asked you to provide the mileage on your car, you might say that it is 45,000 miles, plus or minus 500 miles. The plus or minus amount is the uncertainty in your value. That is, you are indicating that the actual mileage of your car might be as low as 44,500 miles or as high as 45,500 miles, or anywhere in between. All measurements contain some amount of uncertainty. In our example of measuring the length of the paper, we might say that the length of the paper is 11 in., plus or minus 0.2 in. The uncertainty in a measurement, A, is often denoted as δA ("delta A "), so the measurement result would be recorded as $A \pm \delta A$. In our paper example, the length of the paper could be expressed as $11 \text{ in.} \pm 0.2$.

The factors contributing to uncertainty in a measurement include:

1. Limitations of the measuring device,
2. The skill of the person making the measurement,
3. Irregularities in the object being measured,
4. Any other factors that affect the outcome (highly dependent on the situation).

In our example, such factors contributing to the uncertainty could be the following: the smallest division on the ruler is 0.1 in., the person using the ruler has bad eyesight, or one side of the paper is slightly longer than the other. At any rate, the uncertainty in a measurement must be based on a careful consideration of all the factors that might contribute and their possible effects.

Note:
Making Connections: Real-World Connections – Fevers or Chills?
Uncertainty is a critical piece of information, both in physics and in many other real-world applications. Imagine you are caring for a sick child. You suspect the child has a fever, so you check his or her temperature with a thermometer. What if the uncertainty of the thermometer were 3.0°C? If the child's temperature reading was 37.0°C (which is normal body temperature), the "true" temperature could be anywhere from a

hypothermic 34.0^{o}C to a dangerously high 40.0^{o}C. A thermometer with an uncertainty of 3.0^{o}C would be useless.

Percent Uncertainty

One method of expressing uncertainty is as a percent of the measured value. If a measurement A is expressed with uncertainty, δA, the **percent uncertainty** (%unc) is defined to be

Equation:

$$\%\text{ unc} = \frac{\delta A}{A} \times 100\%.$$

Example:

Calculating Percent Uncertainty: A Bag of Apples

A grocery store sells 5-lb bags of apples. You purchase four bags over the course of a month and weigh the apples each time. You obtain the following measurements:

Week 1 weight: 4.8 lb
Week 2 weight: 5.3 lb
Week 3 weight: 4.9 lb
Week 4 weight: 5.4 lb

You determine that the weight of the 5-lb bag has an uncertainty of ± 0.4 lb. What is the percent uncertainty of the bag's weight?

Strategy

First, observe that the expected value of the bag's weight, A, is 5 lb. The uncertainty in this value, δA, is 0.4 lb. We can use the following equation to determine the percent uncertainty of the weight:

Equation:

$$\% \text{ unc} = \frac{\delta A}{A} \times 100\%.$$

Solution
Plug the known values into the equation:
Equation:

$$\% \text{ unc} = \frac{0.4 \text{ lb}}{5 \text{ lb}} \times 100\% = 8\%.$$

Discussion
We can conclude that the weight of the apple bag is $5 \text{ lb} \pm 8\%$. Consider how this percent uncertainty would change if the bag of apples were half as heavy, but the uncertainty in the weight remained the same. Hint for future calculations: when calculating percent uncertainty, always remember that you must multiply the fraction by 100%. If you do not do this, you will have a decimal quantity, not a percent value.

Uncertainties in Calculations

There is an uncertainty in anything calculated from measured quantities. For example, the area of a floor calculated from measurements of its length and width has an uncertainty because the length and width have uncertainties. How big is the uncertainty in something you calculate by multiplication or division? If the measurements going into the calculation have small uncertainties (a few percent or less), then the **method of adding percents** can be used for multiplication or division. This method says that *the percent uncertainty in a quantity calculated by multiplication or division is the sum of the percent uncertainties in the items used to make the calculation*. For example, if a floor has a length of 4.00 m and a width of 3.00 m, with uncertainties of 2% and 1%, respectively, then the area of the floor is 12.0 m^2 and has an uncertainty of 3%. (Expressed as an area this is 0.36 m^2, which we round to 0.4 m^2 since the area of the floor is given to a tenth of a square meter.)
Exercise:
Check Your Understanding

Problem:

A high school track coach has just purchased a new stopwatch. The stopwatch manual states that the stopwatch has an uncertainty of ± 0.05 s. Runners on the track coach's team regularly clock 100-m sprints of 11.49 s to 15.01 s. At the school's last track meet, the first-place sprinter came in at 12.04 s and the second-place sprinter came in at 12.07 s. Will the coach's new stopwatch be helpful in timing the sprint team? Why or why not?

Solution:

No, the uncertainty in the stopwatch is too great to effectively differentiate between the sprint times.

Precision of Measuring Tools and Significant Figures

An important factor in the accuracy and precision of measurements involves the precision of the measuring tool. In general, a precise measuring tool is one that can measure values in very small increments. For example, a standard ruler can measure length to the nearest millimeter, while a caliper can measure length to the nearest 0.01 millimeter. The caliper is a more precise measuring tool because it can measure extremely small differences in length. The more precise the measuring tool, the more precise and accurate the measurements can be.

When we express measured values, we can only list as many digits as we initially measured with our measuring tool. For example, if you use a standard ruler to measure the length of a stick, you may measure it to be 36.7 cm. You could not express this value as 36.71 cm because your measuring tool was not precise enough to measure a hundredth of a centimeter. It should be noted that the last digit in a measured value has been estimated in some way by the person performing the measurement. For example, the person measuring the length of a stick with a ruler notices that the stick length seems to be somewhere in between 36.6 cm and 36.7 cm, and he or she must estimate the value of the last digit. Using the

method of **significant figures**, the rule is that *the last digit written down in a measurement is the first digit with some uncertainty*. In order to determine the number of significant digits in a value, start with the first measured value at the left and count the number of digits through the last digit written on the right. For example, the measured value $36.7\ \text{cm}$ has three digits, or significant figures. Significant figures indicate the precision of a measuring tool that was used to measure a value.

Zeros

Special consideration is given to zeros when counting significant figures. The zeros in 0.053 are not significant, because they are only placekeepers that locate the decimal point. There are two significant figures in 0.053. The zeros in 10.053 are not placekeepers but are significant—this number has five significant figures. The zeros in 1300 may or may not be significant depending on the style of writing numbers. They could mean the number is known to the last digit, or they could be placekeepers. So 1300 could have two, three, or four significant figures. (To avoid this ambiguity, write 1300 in scientific notation.) *Zeros are significant except when they serve only as placekeepers*.

Exercise:

Check Your Understanding

Problem:

Determine the number of significant figures in the following measurements:

a. 0.0009
b. 15,450.0
c. 6×10^3
d. 87.990
e. 30.42

Solution:

(a) 1; the zeros in this number are placekeepers that indicate the decimal point

(b) 6; here, the zeros indicate that a measurement was made to the 0.1 decimal point, so the zeros are significant

(c) 1; the value 10^3 signifies the decimal place, not the number of measured values

(d) 5; the final zero indicates that a measurement was made to the 0.001 decimal point, so it is significant

(e) 4; any zeros located in between significant figures in a number are also significant

Significant Figures in Calculations

When combining measurements with different degrees of accuracy and precision, *the number of significant digits in the final answer can be no greater than the number of significant digits in the least precise measured value*. There are two different rules, one for multiplication and division and the other for addition and subtraction, as discussed below.

1. For multiplication and division: *The result should have the same number of significant figures as the quantity having the least significant figures entering into the calculation*. For example, the area of a circle can be calculated from its radius using $A = \pi r^2$. Let us see how many significant figures the area has if the radius has only two—say, $r = 1.2\text{ m}$. Then,

Equation:

$$A = \pi r^2 = (3.1415927...) \times (1.2\text{ m})^2 = 4.5238934\text{ m}^2$$

is what you would get using a calculator that has an eight-digit output. But because the radius has only two significant figures, it limits the calculated

quantity to two significant figures or
Equation:

$$A=4.5\ \mathrm{m}^2,$$

even though π is good to at least eight digits.

2. For addition and subtraction: *The answer can contain no more decimal places than the least precise measurement.* Suppose that you buy 7.56-kg of potatoes in a grocery store as measured with a scale with precision 0.01 kg. Then you drop off 6.052-kg of potatoes at your laboratory as measured by a scale with precision 0.001 kg. Finally, you go home and add 13.7 kg of potatoes as measured by a bathroom scale with precision 0.1 kg. How many kilograms of potatoes do you now have, and how many significant figures are appropriate in the answer? The mass is found by simple addition and subtraction:
Equation:

$$\begin{array}{ll} & 7.56\ \ \ \mathrm{kg} \\ - & 6.052\ \mathrm{kg} \end{array}\ \ \frac{+13.7\ \ \ \mathrm{kg}}{15.208\ \mathrm{kg}} = 15.2\ \mathrm{kg}.$$

Next, we identify the least precise measurement: 13.7 kg. This measurement is expressed to the 0.1 decimal place, so our final answer must also be expressed to the 0.1 decimal place. Thus, the answer is rounded to the tenths place, giving us 15.2 kg.

Significant Figures in this Text

In this text, most numbers are assumed to have three significant figures. Furthermore, consistent numbers of significant figures are used in all worked examples. You will note that an answer given to three digits is based on input good to at least three digits, for example. If the input has fewer significant figures, the answer will also have fewer significant

figures. Care is also taken that the number of significant figures is reasonable for the situation posed. In some topics, particularly in optics, more accurate numbers are needed and more than three significant figures will be used. Finally, if a number is *exact*, such as the two in the formula for the circumference of a circle, $c = 2\pi r$, it does not affect the number of significant figures in a calculation.

Exercise:

Check Your Understanding

Problem:

Perform the following calculations and express your answer using the correct number of significant digits.

(a) A woman has two bags weighing 13.5 pounds and one bag with a weight of 10.2 pounds. What is the total weight of the bags?

(b) The force F on an object is equal to its mass m multiplied by its acceleration a. If a wagon with mass 55 kg accelerates at a rate of $0.0255\ \text{m/s}^2$, what is the force on the wagon? (The unit of force is called the newton, and it is expressed with the symbol N.)

Solution:

(a) 37.2 pounds; Because the number of bags is an exact value, it is not considered in the significant figures.

(b) 1.4 N; Because the value 55 kg has only two significant figures, the final value must also contain two significant figures.

Note:
PhET Explorations: Estimation
Explore size estimation in one, two, and three dimensions! Multiple levels of difficulty allow for progressive skill improvement.
https://phet.colorado.edu/sims/estimation/estimation_en.html

Summary

- Accuracy of a measured value refers to how close a measurement is to the correct value. The uncertainty in a measurement is an estimate of the amount by which the measurement result may differ from this value.
- Precision of measured values refers to how close the agreement is between repeated measurements.
- The precision of a *measuring tool* is related to the size of its measurement increments. The smaller the measurement increment, the more precise the tool.
- Significant figures express the precision of a measuring tool.
- When multiplying or dividing measured values, the final answer can contain only as many significant figures as the least precise value.
- When adding or subtracting measured values, the final answer cannot contain more decimal places than the least precise value.

Conceptual Questions

Exercise:

Problem:

What is the relationship between the accuracy and uncertainty of a measurement?

Exercise:

Problem:

Prescriptions for vision correction are given in units called *diopters* (D). Determine the meaning of that unit. Obtain information (perhaps by calling an optometrist or performing an internet search) on the minimum uncertainty with which corrections in diopters are determined and the accuracy with which corrective lenses can be produced. Discuss the sources of uncertainties in both the prescription and accuracy in the manufacture of lenses.

Problems & Exercises

Express your answers to problems in this section to the correct number of significant figures and proper units.

Exercise:

Problem:

Suppose that your bathroom scale reads your mass as 65 kg with a 3% uncertainty. What is the uncertainty in your mass (in kilograms)?

Solution:

2 kg

Exercise:

Problem:

A good-quality measuring tape can be off by 0.50 cm over a distance of 20 m. What is its percent uncertainty?

Exercise:

Problem:

(a) A car speedometer has a 5.0% uncertainty. What is the range of possible speeds when it reads 90 km/h? (b) Convert this range to miles per hour. $(1 \text{ km} = 0.6214 \text{ mi})$

Solution:

a. $85.5 \text{ to } 94.5 \text{ km/h}$
b. $53.1 \text{ to } 58.7 \text{ mi/h}$

Exercise:

Problem:

An infant's pulse rate is measured to be 130 ± 5 beats/min. What is the percent uncertainty in this measurement?

Exercise:

Problem:

(a) Suppose that a person has an average heart rate of 72.0 beats/min. How many beats does he or she have in 2.0 y? (b) In 2.00 y? (c) In 2.000 y?

Solution:

(a) 7.6×10^7 beats

(b) 7.57×10^7 beats

(c) 7.57×10^7 beats

Exercise:

Problem:

A can contains 375 mL of soda. How much is left after 308 mL is removed?

Exercise:

Problem:

State how many significant figures are proper in the results of the following calculations: (a) $(106.7)(98.2)/(46.210)(1.01)$ (b) $(18.7)^2$ (c) $\left(1.60 \times 10^{-19}\right)(3712)$.

Solution:

a. 3
b. 3
c. 3

Exercise:

Problem:

(a) How many significant figures are in the numbers 99 and 100? (b) If the uncertainty in each number is 1, what is the percent uncertainty in each? (c) Which is a more meaningful way to express the accuracy of these two numbers, significant figures or percent uncertainties?

Exercise:

Problem:

(a) If your speedometer has an uncertainty of 2.0 km/h at a speed of 90 km/h, what is the percent uncertainty? (b) If it has the same percent uncertainty when it reads 60 km/h, what is the range of speeds you could be going?

Solution:

a) 2.2%

(b) $59 \text{ to } 61 \text{ km/h}$

Exercise:

Problem:

(a) A person's blood pressure is measured to be $120 \pm 2 \text{ mm Hg}$. What is its percent uncertainty? (b) Assuming the same percent uncertainty, what is the uncertainty in a blood pressure measurement of 80 mm Hg?

Exercise:

Problem:

A person measures his or her heart rate by counting the number of beats in 30 s. If 40 ± 1 beats are counted in $30.0 \pm 0.5 \text{ s}$, what is the heart rate and its uncertainty in beats per minute?

Solution:

$80 \pm 3 \text{ beats/min}$

Exercise:

Problem: What is the area of a circle 3.102 cm in diameter?

Exercise:

Problem:

If a marathon runner averages 9.5 mi/h, how long does it take him or her to run a 26.22-mi marathon?

Solution:

2.8 h

Exercise:

Problem:

A marathon runner completes a 42.188-km course in 2 h, 30 min, and 12 s. There is an uncertainty of 25 m in the distance traveled and an uncertainty of 1 s in the elapsed time. (a) Calculate the percent uncertainty in the distance. (b) Calculate the uncertainty in the elapsed time. (c) What is the average speed in meters per second? (d) What is the uncertainty in the average speed?

Exercise:

Problem:

The sides of a small rectangular box are measured to be $1.80 \pm 0.01 \text{ cm}$, $2.05 \pm 0.02 \text{ cm}$, and $3.1 \pm 0.1 \text{ cm}$ long. Calculate its volume and uncertainty in cubic centimeters.

Solution:

$11 \pm 1 \text{ cm}^3$

Exercise:

Problem:

When non-metric units were used in the United Kingdom, a unit of mass called the *pound-mass* (lbm) was employed, where $1 \text{ lbm} = 0.4539 \text{ kg}$. (a) If there is an uncertainty of 0.0001 kg in the pound-mass unit, what is its percent uncertainty? (b) Based on that percent uncertainty, what mass in pound-mass has an uncertainty of 1 kg when converted to kilograms?

Exercise:

Problem:

The length and width of a rectangular room are measured to be $3.955 \pm 0.005 \text{ m}$ and $3.050 \pm 0.005 \text{ m}$. Calculate the area of the room and its uncertainty in square meters.

Solution:

$12.06 \pm 0.04 \text{ m}^2$

Exercise:

Problem:

A car engine moves a piston with a circular cross section of $7.500 \pm 0.002 \text{ cm}$ diameter a distance of $3.250 \pm 0.001 \text{ cm}$ to compress the gas in the cylinder. (a) By what amount is the gas decreased in volume in cubic centimeters? (b) Find the uncertainty in this volume.

Glossary

accuracy
: the degree to which a measured value agrees with correct value for that measurement

method of adding percents

the percent uncertainty in a quantity calculated by multiplication or division is the sum of the percent uncertainties in the items used to make the calculation

percent uncertainty
: the ratio of the uncertainty of a measurement to the measured value, expressed as a percentage

precision
: the degree to which repeated measurements agree with each other

significant figures
: express the precision of a measuring tool used to measure a value

uncertainty
: a quantitative measure of how much your measured values deviate from a standard or expected value

Approximation

- Make reasonable approximations based on given data.

On many occasions, physicists, other scientists, and engineers need to make **approximations** or "guesstimates" for a particular quantity. What is the distance to a certain destination? What is the approximate density of a given item? About how large a current will there be in a circuit? Many approximate numbers are based on formulae in which the input quantities are known only to a limited accuracy. As you develop problem-solving skills (that can be applied to a variety of fields through a study of physics), you will also develop skills at approximating. You will develop these skills through thinking more quantitatively, and by being willing to take risks. As with any endeavor, experience helps, as well as familiarity with units. These approximations allow us to rule out certain scenarios or unrealistic numbers. Approximations also allow us to challenge others and guide us in our approaches to our scientific world. Let us do two examples to illustrate this concept.

Example:
Approximate the Height of a Building
Can you approximate the height of one of the buildings on your campus, or in your neighborhood? Let us make an approximation based upon the height of a person. In this example, we will calculate the height of a 39-story building.
Strategy
Think about the average height of an adult male. We can approximate the height of the building by scaling up from the height of a person.
Solution
Based on information in the example, we know there are 39 stories in the building. If we use the fact that the height of one story is approximately equal to about the length of two adult humans (each human is about 2-m tall), then we can estimate the total height of the building to be
Equation:

$$\frac{2\text{ m}}{1\text{ person}} \times \frac{2\text{ person}}{1\text{ story}} \times 39\text{ stories} = 156\text{ m.}$$

Discussion
You can use known quantities to determine an approximate measurement of unknown quantities. If your hand measures 10 cm across, how many hand lengths equal the width of your desk? What other measurements can you approximate besides length?

Example:
Approximating Vast Numbers: a Trillion Dollars

A bank stack contains one-hundred $100 bills, and is worth $10,000. How many bank stacks make up a trillion dollars? (credit: Andrew Magill)

The U.S. federal deficit in the 2008 fiscal year was a little greater than $10 trillion. Most of us do not have any concept of how much even one trillion actually is. Suppose that you were given a trillion dollars in $100 bills. If you made 100-bill stacks and used them to evenly cover a football field (between the end zones), make an approximation of how high the money pile would become. (We will use feet/inches rather than meters here

because football fields are measured in yards.) One of your friends says 3 in., while another says 10 ft. What do you think?

Strategy

When you imagine the situation, you probably envision thousands of small stacks of 100 wrapped $100 bills, such as you might see in movies or at a bank. Since this is an easy-to-approximate quantity, let us start there. We can find the volume of a stack of 100 bills, find out how many stacks make up one trillion dollars, and then set this volume equal to the area of the football field multiplied by the unknown height.

Solution

(1) Calculate the volume of a stack of 100 bills. The dimensions of a single bill are approximately 3 in. by 6 in. A stack of 100 of these is about 0.5 in. thick. So the total volume of a stack of 100 bills is:

Equation:

$$\begin{aligned}\text{volume of stack} &= \text{length} \times \text{width} \times \text{height},\\ \text{volume of stack} &= 6\text{ in.} \times 3\text{ in.} \times 0.5\text{ in.},\\ \text{volume of stack} &= 9\text{ in.}^3.\end{aligned}$$

(2) Calculate the number of stacks. Note that a trillion dollars is equal to $\$1 \times 10^{12}$, and a stack of one-hundred $100 bills is equal to $\$10{,}000$, or $\$1 \times 10^4$. The number of stacks you will have is:

Equation:

$$\$1 \times 10^{12}(\text{a trillion dollars}) / \ \$1 \times 10^4 \text{ per stack} = 1 \times 10^8 \text{ stacks.}$$

(3) Calculate the area of a football field in square inches. The area of a football field is $100\text{ yd} \times 50\text{ yd}$, which gives $5{,}000\text{ yd}^2$. Because we are working in inches, we need to convert square yards to square inches:

Equation:

$$\begin{gathered}\text{Area} = 5{,}000\text{ yd}^2 \times \frac{3\text{ ft}}{1\text{ yd}} \times \frac{3\text{ ft}}{1\text{ yd}} \times \frac{12\text{ in.}}{1\text{ ft}} \times \frac{12\text{ in.}}{1\text{ ft}} = 6{,}480{,}000\text{ in.}^2,\\ \text{Area} \approx 6 \times 10^6\text{ in.}^2.\end{gathered}$$

This conversion gives us $6 \times 10^6\text{ in.}^2$ for the area of the field. (Note that we are using only one significant figure in these calculations.)

(4) Calculate the total volume of the bills. The volume of all the \$100-bill stacks is $9\text{ in.}^3/\text{stack} \times 10^8\text{ stacks} = 9 \times 10^8\text{ in.}^3$.
(5) Calculate the height. To determine the height of the bills, use the equation:
Equation:

$$\begin{aligned}
\text{volume of bills} &= \text{area of field} \times \text{height of money:} \\
\text{Height of money} &= \tfrac{\text{volume of bills}}{\text{area of field}}, \\
\text{Height of money} &= \tfrac{9\times10^8\text{in.}^3}{6\times10^6\text{in.}^2} = 1.33 \times 10^2\text{in.}, \\
\text{Height of money} &\approx 1 \times 10^2\text{in.} = 100\text{ in.}
\end{aligned}$$

The height of the money will be about 100 in. high. Converting this value to feet gives
Equation:

$$100\text{ in.} \times \frac{1\text{ ft}}{12\text{ in.}} = 8.33\text{ ft} \approx 8\text{ ft}.$$

Discussion
The final approximate value is much higher than the early estimate of 3 in., but the other early estimate of 10 ft (120 in.) was roughly correct. How did the approximation measure up to your first guess? What can this exercise tell you in terms of rough “guesstimates” versus carefully calculated approximations?

Exercise:
Check Your Understanding

Problem:

Using mental math and your understanding of fundamental units, approximate the area of a regulation basketball court. Describe the process you used to arrive at your final approximation.

Solution:

An average male is about two meters tall. It would take approximately 15 men laid out end to end to cover the length, and about 7 to cover the width. That gives an approximate area of $420\ \text{m}^2$.

Summary

Scientists often approximate the values of quantities to perform calculations and analyze systems.

Problems & Exercises

Exercise:

Problem: How many heartbeats are there in a lifetime?

Solution:

Sample answer: 2×10^9 heartbeats

Exercise:

Problem:

A generation is about one-third of a lifetime. Approximately how many generations have passed since the year 0 AD?

Exercise:

Problem:

How many times longer than the mean life of an extremely unstable atomic nucleus is the lifetime of a human? (Hint: The lifetime of an unstable atomic nucleus is on the order of 10^{-22} s.)

Solution:

Sample answer: 2×10^{31} if an average human lifetime is taken to be about 70 years.

Exercise:

Problem:

Calculate the approximate number of atoms in a bacterium. Assume that the average mass of an atom in the bacterium is ten times the mass of a hydrogen atom. (Hint: The mass of a hydrogen atom is on the order of 10^{-27} kg and the mass of a bacterium is on the order of 10^{-15} kg.)

This color-enhanced photo shows *Salmonella typhimurium* (red) attacking human cells. These bacteria are commonly known for causing foodborne illness. Can you estimate the number of atoms in each bacterium? (credit: Rocky Mountain Laboratories, NIAID, NIH)

Exercise:

Problem:

Approximately how many atoms thick is a cell membrane, assuming all atoms there average about twice the size of a hydrogen atom?

Solution:

Sample answer: 50 atoms

Exercise:

Problem:

(a) What fraction of Earth's diameter is the greatest ocean depth? (b) The greatest mountain height?

Exercise:

Problem:

(a) Calculate the number of cells in a hummingbird assuming the mass of an average cell is ten times the mass of a bacterium. (b) Making the same assumption, how many cells are there in a human?

Solution:

Sample answers:

(a) 10^{12} cells/hummingbird

(b) 10^{16} cells/human

Exercise:

Problem:

Assuming one nerve impulse must end before another can begin, what is the maximum firing rate of a nerve in impulses per second?

Glossary

approximation
: an estimated value based on prior experience and reasoning

# Introduction to the Physics of Hearing
class="introduction"

This tree fell some time ago. When it fell, atoms in the air were disturbed. Physicists would call this disturbance sound whether someone was around to hear it or not. (credit: B.A. Bowen Photography)

If a tree falls in the forest and no one is there to hear it, does it make a sound? The answer to this old philosophical question depends on how you define sound. If sound only exists when someone is around to perceive it, then there was no sound. However, if we define sound in terms of physics; that is, a disturbance of the atoms in matter transmitted from its origin outward (in other words, a wave), then there *was* a sound, even if nobody was around to hear it.

Such a wave is the physical phenomenon we call *sound.* Its perception is hearing. Both the physical phenomenon and its perception are interesting and will be considered in this text. We shall explore both sound and hearing; they are related, but are not the same thing. We will also explore the many practical uses of sound waves, such as in medical imaging.

Sound

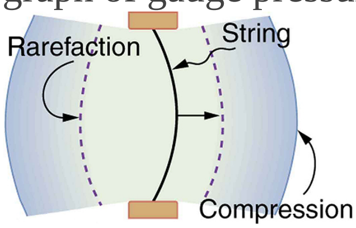
- Define sound and hearing.
- Describe sound as a longitudinal wave.

This glass has been shattered by a high-intensity sound wave of the same frequency as the resonant frequency of the glass. While the sound is not visible, the effects of the sound prove its existence. (credit: ||read||, Flickr)

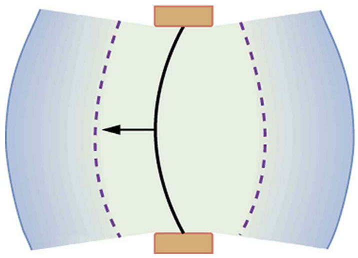
Sound can be used as a familiar illustration of waves. Because hearing is one of our most important senses, it is interesting to see how the physical properties of sound correspond to our perceptions of it. **Hearing** is the perception of sound, just as vision is the perception of visible light. But sound has important applications beyond hearing. Ultrasound, for example, is not heard but can be employed to form medical images and is also used in treatment.

The physical phenomenon of **sound** is defined to be a disturbance of matter that is transmitted from its source outward. Sound is a wave. On the atomic scale, it is a disturbance of atoms that is far more ordered than their thermal motions. In many instances, sound is a periodic wave, and the atoms undergo simple harmonic motion. In this text, we shall explore such periodic sound waves.

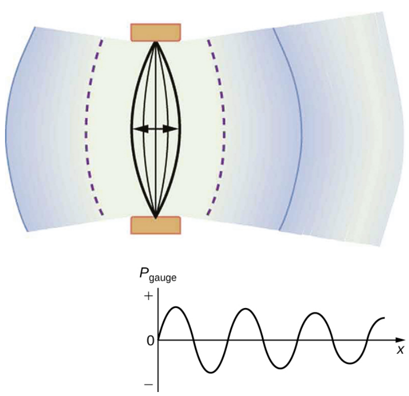
A vibrating string produces a sound wave as illustrated in [link], [link], and [link]. As the string oscillates back and forth, it transfers energy to the air, mostly as thermal energy created by turbulence. But a small part of the string's energy goes into compressing and expanding the surrounding air, creating slightly higher and lower local pressures. These compressions (high pressure regions) and rarefactions (low pressure regions) move out as longitudinal pressure waves having the same frequency as the string—they are the disturbance that is a sound wave. (Sound waves in air and most fluids are longitudinal, because fluids have almost no shear strength. In solids, sound waves can be both transverse and longitudinal.) [link] shows a graph of gauge pressure versus distance from the vibrating string.



A vibrating string moving to the right compresses the air in front of it and expands the air behind it.



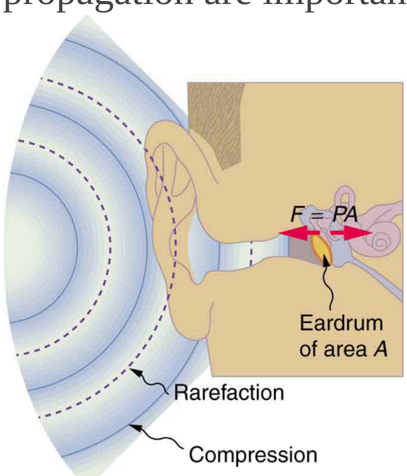
As the string moves to the left, it creates another compression and rarefaction as the ones on the right move away from the string.



After many vibrations, there are a series of compressions and rarefactions moving out from the string as a sound wave. The graph shows gauge pressure versus

distance from the source. Pressures vary only slightly from atmospheric for ordinary sounds.

The amplitude of a sound wave decreases with distance from its source, because the energy of the wave is spread over a larger and larger area. But it is also absorbed by objects, such as the eardrum in [link], and converted to thermal energy by the viscosity of air. In addition, during each compression a little heat transfers to the air and during each rarefaction even less heat transfers from the air, so that the heat transfer reduces the organized disturbance into random thermal motions. (These processes can be viewed as a manifestation of the second law of thermodynamics presented in Introduction to the Second Law of Thermodynamics: Heat Engines and Their Efficiency.) Whether the heat transfer from compression to rarefaction is significant depends on how far apart they are—that is, it depends on wavelength. Wavelength, frequency, amplitude, and speed of propagation are important for sound, as they are for all waves.



Sound wave compressions and rarefactions travel up the ear canal and

force the eardrum to vibrate. There is a net force on the eardrum, since the sound wave pressures differ from the atmospheric pressure found behind the eardrum. A complicated mechanism converts the vibrations to nerve impulses, which are perceived by the person.

Note:
PhET Explorations: Wave Interference
WMake waves with a dripping faucet, audio speaker, or laser! Add a second source or a pair of slits to create an interference pattern.
https://archive.cnx.org/specials/2fe7ad15-b00e-4402-b068-ff503985a18f/wave-interference/

Section Summary

- Sound is a disturbance of matter that is transmitted from its source outward.
- Sound is one type of wave.

- Hearing is the perception of sound.

Glossary

sound
: a disturbance of matter that is transmitted from its source outward

hearing
: the perception of sound

Speed of Sound, Frequency, and Wavelength

- Define pitch.
- Describe the relationship between the speed of sound, its frequency, and its wavelength.
- Describe the effects on the speed of sound as it travels through various media.
- Describe the effects of temperature on the speed of sound.

When a firework explodes, the light energy is perceived before the sound energy. Sound travels more slowly than light does. (credit: Dominic Alves, Flickr)

Sound, like all waves, travels at a certain speed and has the properties of frequency and wavelength. You can observe direct evidence of the speed of sound while watching a fireworks display. The flash of an explosion is seen well before its sound is heard, implying both that sound travels at a finite speed and that it is much slower than light. You can also directly sense the frequency of a sound. Perception of frequency is called **pitch**. The wavelength of sound is not directly sensed, but indirect evidence is found in the correlation of the size of musical instruments with their pitch. Small

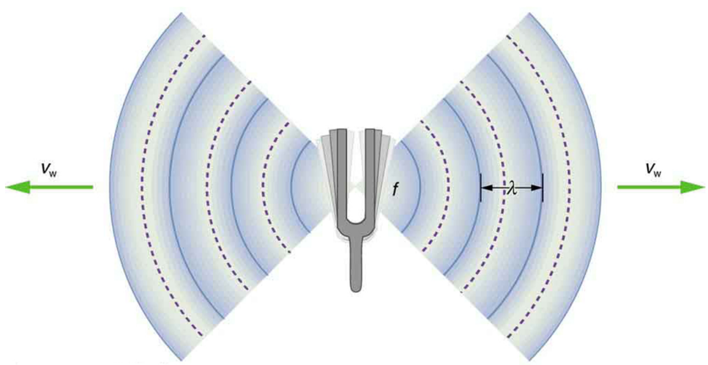
instruments, such as a piccolo, typically make high-pitch sounds, while large instruments, such as a tuba, typically make low-pitch sounds. High pitch means small wavelength, and the size of a musical instrument is directly related to the wavelengths of sound it produces. So a small instrument creates short-wavelength sounds. Similar arguments hold that a large instrument creates long-wavelength sounds.

The relationship of the speed of sound, its frequency, and wavelength is the same as for all waves:

Equation:

$$v_{\text{w}} = f\lambda,$$

where v_{w} is the speed of sound, f is its frequency, and λ is its wavelength. The wavelength of a sound is the distance between adjacent identical parts of a wave—for example, between adjacent compressions as illustrated in [link]. The frequency is the same as that of the source and is the number of waves that pass a point per unit time.



A sound wave emanates from a source vibrating at a frequency f, propagates at v_{w}, and has a wavelength λ.

[link] makes it apparent that the speed of sound varies greatly in different media. The speed of sound in a medium is determined by a combination of the medium’s rigidity (or compressibility in gases) and its density. The

more rigid (or less compressible) the medium, the faster the speed of sound. This observation is analogous to the fact that the frequency of a simple harmonic motion is directly proportional to the stiffness of the oscillating object. The greater the density of a medium, the slower the speed of sound. This observation is analogous to the fact that the frequency of a simple harmonic motion is inversely proportional to the mass of the oscillating object. The speed of sound in air is low, because air is compressible. Because liquids and solids are relatively rigid and very difficult to compress, the speed of sound in such media is generally greater than in gases.

Medium	v_w**(m/s)**
Gases at $0°C$	
Air	331
Carbon dioxide	259
Oxygen	316
Helium	965
Hydrogen	1290
Liquids at $20°C$	
Ethanol	1160
Mercury	1450
Water, fresh	1480

Medium	v_w**(m/s)**
Sea water	1540
Human tissue	1540
Solids (longitudinal or bulk)	
Vulcanized rubber	54
Polyethylene	920
Marble	3810
Glass, Pyrex	5640
Lead	1960
Aluminum	5120
Steel	5960

Speed of Sound in Various Media

Earthquakes, essentially sound waves in Earth's crust, are an interesting example of how the speed of sound depends on the rigidity of the medium. Earthquakes have both longitudinal and transverse components, and these travel at different speeds. The bulk modulus of granite is greater than its shear modulus. For that reason, the speed of longitudinal or pressure waves (P-waves) in earthquakes in granite is significantly higher than the speed of transverse or shear waves (S-waves). Both components of earthquakes travel slower in less rigid material, such as sediments. P-waves have speeds of 4 to 7 km/s, and S-waves correspondingly range in speed from 2 to 5 km/s, both being faster in more rigid material. The P-wave gets progressively farther ahead of the S-wave as they travel through Earth's crust. The time between the P- and S-waves is routinely used to determine the distance to their source, the epicenter of the earthquake.

The speed of sound is affected by temperature in a given medium. For air at sea level, the speed of sound is given by

Equation:

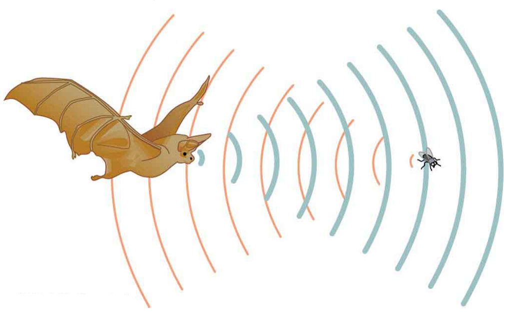
$$v_{\mathrm{w}} = (331\ \mathrm{m/s})\sqrt{\frac{T}{273\ \mathrm{K}}},$$

where the temperature (denoted as T) is in units of kelvin. The speed of sound in gases is related to the average speed of particles in the gas, v_{rms}, and that

Equation:

$$v_{\mathrm{rms}} = \sqrt{\frac{3\,kT}{m}},$$

where k is the Boltzmann constant ($1.38 \times 10^{-23}\ \mathrm{J/K}$) and m is the mass of each (identical) particle in the gas. So, it is reasonable that the speed of sound in air and other gases should depend on the square root of temperature. While not negligible, this is not a strong dependence. At $0^{\mathrm{o}}\mathrm{C}$, the speed of sound is 331 m/s, whereas at $20.0^{\mathrm{o}}\mathrm{C}$ it is 343 m/s, less than a 4% increase. [link] shows a use of the speed of sound by a bat to sense distances. Echoes are also used in medical imaging.



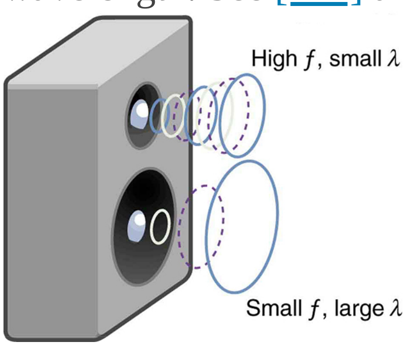
A bat uses sound echoes to find its way about and to catch prey. The time for the echo to return is directly proportional to the distance.

One of the more important properties of sound is that its speed is nearly independent of frequency. This independence is certainly true in open air for sounds in the audible range of 20 to 20,000 Hz. If this independence were not true, you would certainly notice it for music played by a marching band in a football stadium, for example. Suppose that high-frequency sounds traveled faster—then the farther you were from the band, the more the sound from the low-pitch instruments would lag that from the high-pitch ones. But the music from all instruments arrives in cadence independent of distance, and so all frequencies must travel at nearly the same speed. Recall that

Equation:

$$v_{\text{w}} = f\lambda.$$

In a given medium under fixed conditions, v_{w} is constant, so that there is a relationship between f and λ; the higher the frequency, the smaller the wavelength. See [link] and consider the following example.



Because they travel at the same speed in a given medium, low-frequency sounds must have a greater wavelength than high-frequency sounds.

Here, the lower-frequency sounds are emitted by the large speaker, called a woofer, while the higher-frequency sounds are emitted by the small speaker, called a tweeter.

Example:
Calculating Wavelengths: What Are the Wavelengths of Audible Sounds?

Calculate the wavelengths of sounds at the extremes of the audible range, 20 and 20,000 Hz, in 30.0°C air. (Assume that the frequency values are accurate to two significant figures.)

Strategy

To find wavelength from frequency, we can use $v_{\mathrm{w}} = f\lambda$.

Solution

1. Identify knowns. The value for v_{w}, is given by
   Equation:
   $$v_{\mathrm{w}} = (331\ \mathrm{m/s})\sqrt{\frac{T}{273\ \mathrm{K}}}.$$
2. Convert the temperature into kelvin and then enter the temperature into the equation
   Equation:
   $$v_{\mathrm{w}} = (331\ \mathrm{m/s})\sqrt{\frac{303\ \mathrm{K}}{273\ \mathrm{K}}} = 348.7\ \mathrm{m/s}.$$

3. Solve the relationship between speed and wavelength for λ:
   Equation:

$$\lambda = \frac{v_{\mathrm{w}}}{f}.$$

4. Enter the speed and the minimum frequency to give the maximum wavelength:
   Equation:

$$\lambda_{\mathrm{max}} = \frac{348.7\ \mathrm{m/s}}{20\ \mathrm{Hz}} = 17\ \mathrm{m}.$$

5. Enter the speed and the maximum frequency to give the minimum wavelength:
   Equation:

$$\lambda_{\mathrm{min}} = \frac{348.7\ \mathrm{m/s}}{20{,}000\ \mathrm{Hz}} = 0.017\ \mathrm{m} = 1.7\ \mathrm{cm}.$$

Discussion
Because the product of f multiplied by λ equals a constant, the smaller f is, the larger λ must be, and vice versa.

The speed of sound can change when sound travels from one medium to another. However, the frequency usually remains the same because it is like a driven oscillation and has the frequency of the original source. If v_{w} changes and f remains the same, then the wavelength λ must change. That is, because $v_{\mathrm{w}} = f\lambda$, the higher the speed of a sound, the greater its wavelength for a given frequency.

Note:
Making Connections: Take-Home Investigation—Voice as a Sound Wave

Suspend a sheet of paper so that the top edge of the paper is fixed and the bottom edge is free to move. You could tape the top edge of the paper to the edge of a table. Gently blow near the edge of the bottom of the sheet and note how the sheet moves. Speak softly and then louder such that the sounds hit the edge of the bottom of the paper, and note how the sheet moves. Explain the effects.

Exercise:
Check Your Understanding

Problem:

Imagine you observe two fireworks explode. You hear the explosion of one as soon as you see it. However, you see the other firework for several milliseconds before you hear the explosion. Explain why this is so.

Solution:

Sound and light both travel at definite speeds. The speed of sound is slower than the speed of light. The first firework is probably very close by, so the speed difference is not noticeable. The second firework is farther away, so the light arrives at your eyes noticeably sooner than the sound wave arrives at your ears.

Exercise:
Check Your Understanding

Problem:

You observe two musical instruments that you cannot identify. One plays high-pitch sounds and the other plays low-pitch sounds. How could you determine which is which without hearing either of them play?

Solution:

Compare their sizes. High-pitch instruments are generally smaller than low-pitch instruments because they generate a smaller wavelength.

Section Summary

The relationship of the speed of sound v_{w}, its frequency f, and its wavelength λ is given by
Equation:

$$v_{\text{w}} = f\lambda,$$

which is the same relationship given for all waves.

In air, the speed of sound is related to air temperature T by
Equation:

$$v_{\text{w}} = (331\ \text{m/s})\sqrt{\frac{T}{273\ \text{K}}}.$$

v_{w} is the same for all frequencies and wavelengths.

Conceptual Questions

Exercise:

Problem:

How do sound vibrations of atoms differ from thermal motion?

Exercise:

Problem:

When sound passes from one medium to another where its propagation speed is different, does its frequency or wavelength change? Explain your answer briefly.

Problems & Exercises

Exercise:

Problem:

When poked by a spear, an operatic soprano lets out a 1200-Hz shriek. What is its wavelength if the speed of sound is 345 m/s?

Solution:

0.288 m

Exercise:

Problem:

What frequency sound has a 0.10-m wavelength when the speed of sound is 340 m/s?

Exercise:

Problem:

Calculate the speed of sound on a day when a 1500 Hz frequency has a wavelength of 0.221 m.

Solution:

332 m/s

Exercise:

Problem:

(a) What is the speed of sound in a medium where a 100-kHz frequency produces a 5.96-cm wavelength? (b) Which substance in [link] is this likely to be?

Exercise:

Problem:

Show that the speed of sound in $20.0^{\circ}C$ air is 343 m/s, as claimed in the text.

Solution:
Equation:

$$\begin{aligned} v_{\text{w}} &= (331\text{ m/s})\sqrt{\frac{T}{273\text{ K}}} = (331\text{ m/s})\sqrt{\frac{293\text{ K}}{273\text{ K}}} \\ &= 343\text{ m/s} \end{aligned}$$

Exercise:

Problem:

Air temperature in the Sahara Desert can reach $56.0^{\circ}C$ (about $134^{\circ}F$). What is the speed of sound in air at that temperature?

Exercise:

Problem:

Dolphins make sounds in air and water. What is the ratio of the wavelength of a sound in air to its wavelength in seawater? Assume air temperature is $20.0^{\circ}C$.

Solution:

0.223

Exercise:

Problem:

A sonar echo returns to a submarine 1.20 s after being emitted. What is the distance to the object creating the echo? (Assume that the submarine is in the ocean, not in fresh water.)

Exercise:

Problem:

(a) If a submarine's sonar can measure echo times with a precision of 0.0100 s, what is the smallest difference in distances it can detect? (Assume that the submarine is in the ocean, not in fresh water.)

(b) Discuss the limits this time resolution imposes on the ability of the sonar system to detect the size and shape of the object creating the echo.

Solution:

(a) 7.70 m

(b) This means that sonar is good for spotting and locating large objects, but it isn't able to resolve smaller objects, or detect the detailed shapes of objects. Objects like ships or large pieces of airplanes can be found by sonar, while smaller pieces must be found by other means.

Exercise:

Problem:

A physicist at a fireworks display times the lag between seeing an explosion and hearing its sound, and finds it to be 0.400 s. (a) How far away is the explosion if air temperature is 24.0°C and if you neglect the time taken for light to reach the physicist? (b) Calculate the distance to the explosion taking the speed of light into account. Note that this distance is negligibly greater.

Exercise:

Problem:

Suppose a bat uses sound echoes to locate its insect prey, 3.00 m away. (See [link].) (a) Calculate the echo times for temperatures of 5.00°C and 35.0°C. (b) What percent uncertainty does this cause for the bat in locating the insect? (c) Discuss the significance of this uncertainty and whether it could cause difficulties for the bat. (In practice, the bat continues to use sound as it closes in, eliminating most of any difficulties imposed by this and other effects, such as motion of the prey.)

Solution:

(a) 18.0 ms, 17.1 ms

(b) 5.00%

(c) This uncertainty could definitely cause difficulties for the bat, if it didn’t continue to use sound as it closed in on its prey. A 5% uncertainty could be the difference between catching the prey around the neck or around the chest, which means that it could miss grabbing its prey.

Glossary

pitch
: the perception of the frequency of a sound

Sound Intensity and Sound Level

- Define intensity, sound intensity, and sound pressure level.
- Calculate sound intensity levels in decibels (dB).

Noise on crowded roadways like this one in Delhi makes it hard to hear others unless they shout. (credit: Lingaraj G J, Flickr)

In a quiet forest, you can sometimes hear a single leaf fall to the ground. After settling into bed, you may hear your blood pulsing through your ears. But when a passing motorist has his stereo turned up, you cannot even hear what the person next to you in your car is saying. We are all very familiar with the loudness of sounds and aware that they are related to how energetically the source is vibrating. In cartoons depicting a screaming person (or an animal making a loud noise), the cartoonist often shows an open mouth with a vibrating uvula, the hanging tissue at the back of the mouth, to suggest a loud sound coming from the throat [link]. High noise exposure is hazardous to hearing, and it is common for musicians to have hearing losses that are sufficiently severe that they interfere with the musicians' abilities to perform. The relevant physical quantity is sound intensity, a concept that is valid for all sounds whether or not they are in the audible range.

Intensity is defined to be the power per unit area carried by a wave. Power is the rate at which energy is transferred by the wave. In equation form, **intensity** I is

Equation:

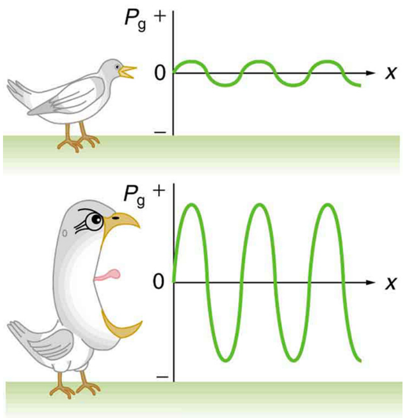
$$I = \frac{P}{A},$$

where P is the power through an area A. The SI unit for I is W/m^2. The intensity of a sound wave is related to its amplitude squared by the following relationship:

Equation:

$$I = \frac{(\Delta p)^2}{2\rho v_{\text{w}}}.$$

Here Δp is the pressure variation or pressure amplitude (half the difference between the maximum and minimum pressure in the sound wave) in units of pascals (Pa) or N/m^2. (We are using a lower case p for pressure to distinguish it from power, denoted by P above.) The energy (as kinetic energy $\frac{mv^2}{2}$) of an oscillating element of air due to a traveling sound wave is proportional to its amplitude squared. In this equation, ρ is the density of the material in which the sound wave travels, in units of kg/m^3, and v_{w} is the speed of sound in the medium, in units of m/s. The pressure variation is proportional to the amplitude of the oscillation, and so I varies as $(\Delta p)^2$ ([link]). This relationship is consistent with the fact that the sound wave is produced by some vibration; the greater its pressure amplitude, the more the air is compressed in the sound it creates.



Graphs of the gauge pressures in two sound waves of different intensities. The more intense sound is produced by a source that has larger-amplitude oscillations and has greater pressure maxima and minima. Because pressures are higher in the greater-intensity sound, it can exert larger forces on the objects it encounters.

Sound intensity levels are quoted in decibels (dB) much more often than sound intensities in watts per meter squared. Decibels are the unit of choice in the scientific literature as well as in the popular media. The reasons for this choice of units are related to how we perceive sounds. How our ears perceive sound can be more accurately described by the logarithm of the

intensity rather than directly to the intensity. The **sound intensity level** β in decibels of a sound having an intensity I in watts per meter squared is defined to be

Equation:

$$\beta\,(\text{dB}) = 10\,\log_{10}\left(\frac{I}{I_0}\right),$$

where $I_0 = 10^{-12}\,\text{W/m}^2$ is a reference intensity. In particular, I_0 is the lowest or threshold intensity of sound a person with normal hearing can perceive at a frequency of 1000 Hz. Sound intensity level is not the same as intensity. Because β is defined in terms of a ratio, it is a unitless quantity telling you the *level* of the sound relative to a fixed standard ($10^{-12}\,\text{W/m}^2$, in this case). The units of decibels (dB) are used to indicate this ratio is multiplied by 10 in its definition. The bel, upon which the decibel is based, is named for Alexander Graham Bell, the inventor of the telephone.

Sound intensity level *β* (dB)	**Intensity *I*(W/m^2)**	**Example/effect**
0	1×10^{-12}	Threshold of hearing at 1000 Hz
10	1×10^{-11}	Rustle of leaves
20	1×10^{-10}	Whisper at 1 m distance
30	1×10^{-9}	Quiet home

Sound intensity level β (dB)	Intensity $I(\mathrm{W/m^2})$	Example/effect
40	1×10^{-8}	Average home
50	1×10^{-7}	Average office, soft music
60	1×10^{-6}	Normal conversation
70	1×10^{-5}	Noisy office, busy traffic
80	1×10^{-4}	Loud radio, classroom lecture
90	1×10^{-3}	Inside a heavy truck; damage from prolonged exposure[footnote] Several government agencies and health-related professional associations recommend that 85 dB not be exceeded for 8-hour daily exposures in the absence of hearing protection.
100	1×10^{-2}	Noisy factory, siren at 30 m; damage from 8 h per day exposure
110	1×10^{-1}	Damage from 30 min per day exposure
120	1	Loud rock concert, pneumatic chipper at 2 m; threshold of pain
140	1×10^{2}	Jet airplane at 30 m; severe pain, damage in seconds
160	1×10^{4}	Bursting of eardrums

Sound Intensity Levels and Intensities

Sound Intensity Levels and Intensities

The decibel level of a sound having the threshold intensity of $10^{-12}\ \text{W/m}^2$ is $\beta = 0\ \text{dB}$, because $\log_{10} 1 = 0$. That is, the threshold of hearing is 0 decibels. [link] gives levels in decibels and intensities in watts per meter squared for some familiar sounds.

One of the more striking things about the intensities in [link] is that the intensity in watts per meter squared is quite small for most sounds. The ear is sensitive to as little as a trillionth of a watt per meter squared—even more impressive when you realize that the area of the eardrum is only about $1\ \text{cm}^2$, so that only 10^{-16} W falls on it at the threshold of hearing! Air molecules in a sound wave of this intensity vibrate over a distance of less than one molecular diameter, and the gauge pressures involved are less than 10^{-9} atm.

Another impressive feature of the sounds in [link] is their numerical range. Sound intensity varies by a factor of 10^{12} from threshold to a sound that causes damage in seconds. You are unaware of this tremendous range in sound intensity because how your ears respond can be described approximately as the logarithm of intensity. Thus, sound intensity levels in decibels fit your experience better than intensities in watts per meter squared. The decibel scale is also easier to relate to because most people are more accustomed to dealing with numbers such as 0, 53, or 120 than numbers such as 1.00×10^{-11}.

One more observation readily verified by examining [link] or using $I = \frac{(\Delta p)}{2\rho v_{\text{w}}}^2$ is that each factor of 10 in intensity corresponds to 10 dB. For example, a 90 dB sound compared with a 60 dB sound is 30 dB greater, or three factors of 10 (that is, 10^3 times) as intense. Another example is that if one sound is 10^7 as intense as another, it is 70 dB higher. See [link].

I_2/I_1	$\beta_2 - \beta_1$
2.0	3.0 dB
5.0	7.0 dB
10.0	10.0 dB

Ratios of Intensities and Corresponding Differences in Sound Intensity Levels

Example:
Calculating Sound Intensity Levels: Sound Waves
Calculate the sound intensity level in decibels for a sound wave traveling in air at 0ºC and having a pressure amplitude of 0.656 Pa.
Strategy
We are given Δp, so we can calculate I using the equation $I = (\Delta p)^2/(2pv_\text{w})^2$. Using I, we can calculate β straight from its definition in $\beta\ (\text{dB}) = 10\ \log_{10}(I/I_0)$.
Solution
(1) Identify knowns:
Sound travels at 331 m/s in air at 0ºC.
Air has a density of $1.29\ \text{kg/m}^3$ at atmospheric pressure and 0ºC.
(2) Enter these values and the pressure amplitude into $I = (\Delta p)^2/\ (2\rho v_\text{w})$:
Equation:

$$I = \frac{(\Delta p)^2}{2\rho v_\text{w}} = \frac{(0.656\ \text{Pa})^2}{2\left(1.29\ \text{kg/m}^3\right)(331\ \text{m/s})} = 5.04 \times 10^{-4}\ \text{W/m}^2.$$

(3) Enter the value for I and the known value for I_0 into $\beta\ (\text{dB}) = 10\ \log_{10}(I/I_0)$. Calculate to find the sound intensity level in decibels:
Equation:

$$10\ \log_{10}\left(5.04 \times 10^{8}\right) = 10\ \left(8.70\right)\ \text{dB} = 87\ \text{dB}.$$

Discussion
This 87 dB sound has an intensity five times as great as an 80 dB sound. So a factor of five in intensity corresponds to a difference of 7 dB in sound intensity level. This value is true for any intensities differing by a factor of five.

Example:
Change Intensity Levels of a Sound: What Happens to the Decibel Level?
Show that if one sound is twice as intense as another, it has a sound level about 3 dB higher.
Strategy
You are given that the ratio of two intensities is 2 to 1, and are then asked to find the difference in their sound levels in decibels. You can solve this problem using of the properties of logarithms.
Solution
(1) Identify knowns:
The ratio of the two intensities is 2 to 1, or:
Equation:

$$\frac{I_2}{I_1} = 2.00.$$

We wish to show that the difference in sound levels is about 3 dB. That is, we want to show:
Equation:

$$\beta_2 - \beta_1 = 3\ \text{dB}.$$

Note that:
Equation:

$$\log_{10} b - \log_{10} a = \log_{10}\left(\frac{b}{a}\right).$$

(2) Use the definition of β to get:

Equation:

$$\beta_2 - \beta_1 = 10 \log_{10}\left(\frac{I_2}{I_1}\right) = 10 \log_{10} 2.00 = 10\,(0.301)\text{ dB.}$$

Thus,

Equation:

$$\beta_2 - \beta_1 = 3.01\text{ dB.}$$

Discussion

This means that the two sound intensity levels differ by 3.01 dB, or about 3 dB, as advertised. Note that because only the ratio I_2/I_1 is given (and not the actual intensities), this result is true for any intensities that differ by a factor of two. For example, a 56.0 dB sound is twice as intense as a 53.0 dB sound, a 97.0 dB sound is half as intense as a 100 dB sound, and so on.

It should be noted at this point that there is another decibel scale in use, called the **sound pressure level**, based on the ratio of the pressure amplitude to a reference pressure. This scale is used particularly in applications where sound travels in water. It is beyond the scope of most introductory texts to treat this scale because it is not commonly used for sounds in air, but it is important to note that very different decibel levels may be encountered when sound pressure levels are quoted. For example, ocean noise pollution produced by ships may be as great as 200 dB expressed in the sound pressure level, where the more familiar sound intensity level we use here would be something under 140 dB for the same sound.

Note:

Take-Home Investigation: Feeling Sound

Find a CD player and a CD that has rock music. Place the player on a light table, insert the CD into the player, and start playing the CD. Place your hand gently on the table next to the speakers. Increase the volume and note the level when the table just begins to vibrate as the rock music plays. Increase the reading on the volume control until it doubles. What has happened to the vibrations?

Exercise:
Check Your Understanding

Problem:

Describe how amplitude is related to the loudness of a sound.

Solution:

Amplitude is directly proportional to the experience of loudness. As amplitude increases, loudness increases.

Exercise:
Check Your Understanding

Problem:

Identify common sounds at the levels of 10 dB, 50 dB, and 100 dB.

Solution:

10 dB: Running fingers through your hair.

50 dB: Inside a quiet home with no television or radio.

100 dB: Take-off of a jet plane.

Section Summary

- Intensity is the same for a sound wave as was defined for all waves; it is
  Equation:
  $$I = \frac{P}{A},$$
  where P is the power crossing area A. The SI unit for I is watts per meter squared. The intensity of a sound wave is also related to the pressure amplitude Δp
  Equation:
  $$I = \frac{(\Delta p)^2}{2\rho v_{\mathrm{w}}},$$
  where ρ is the density of the medium in which the sound wave travels and v_{w} is the speed of sound in the medium.
- Sound intensity level in units of decibels (dB) is
  Equation:
  $$\beta\,(\mathrm{dB}) = 10\,\log_{10}\left(\frac{I}{I_0}\right),$$
  where $I_0 = 10^{-12}\ \mathrm{W/m^2}$ is the threshold intensity of hearing.

Conceptual Questions

Exercise:

Problem:

Six members of a synchronized swim team wear earplugs to protect themselves against water pressure at depths, but they can still hear the music and perform the combinations in the water perfectly. One day, they were asked to leave the pool so the dive team could practice a few dives, and they tried to practice on a mat, but seemed to have a lot more difficulty. Why might this be?

Exercise:

Problem:

A community is concerned about a plan to bring train service to their downtown from the town's outskirts. The current sound intensity level, even though the rail yard is blocks away, is 70 dB downtown. The mayor assures the public that there will be a difference of only 30 dB in sound in the downtown area. Should the townspeople be concerned? Why?

Problems & Exercises

Exercise:

Problem:

What is the intensity in watts per meter squared of 85.0-dB sound?

Solution:
Equation:

$$3.16 \times 10^{-4}\ \mathrm{W/m^2}$$

Exercise:

Problem:

The warning tag on a lawn mower states that it produces noise at a level of 91.0 dB. What is this in watts per meter squared?

Exercise:

Problem:

A sound wave traveling in $20^{\circ}\mathrm{C}$ air has a pressure amplitude of 0.5 Pa. What is the intensity of the wave?

Solution:
Equation:

$$3.04 \times 10^{-4}\ \mathrm{W/m^2}$$

Exercise:

Problem:

What intensity level does the sound in the preceding problem correspond to?

Exercise:

Problem:

What sound intensity level in dB is produced by earphones that create an intensity of $4.00 \times 10^{-2}\ \mathrm{W/m^2}$?

Solution:

106 dB

Exercise:

Problem:

Show that an intensity of $10^{-12}\ \mathrm{W/m^2}$ is the same as $10^{-16}\ \mathrm{W/cm^2}$.

Exercise:

Problem:

(a) What is the decibel level of a sound that is twice as intense as a 90.0-dB sound? (b) What is the decibel level of a sound that is one-fifth as intense as a 90.0-dB sound?

Solution:

(a) 93 dB

(b) 83 dB

Exercise:

Problem:

(a) What is the intensity of a sound that has a level 7.00 dB lower than a $4.00 \times 10^{-9}\ \text{W/m}^2$ sound? (b) What is the intensity of a sound that is 3.00 dB higher than a $4.00 \times 10^{-9}\ \text{W/m}^2$ sound?

Exercise:

Problem:

(a) How much more intense is a sound that has a level 17.0 dB higher than another? (b) If one sound has a level 23.0 dB less than another, what is the ratio of their intensities?

Solution:

(a) 50.1

(b) 5.01×10^{-3} or $\frac{1}{200}$

Exercise:

Problem:

People with good hearing can perceive sounds as low in level as –8.00 dB at a frequency of 3000 Hz. What is the intensity of this sound in watts per meter squared?

Exercise:

Problem:

If a large housefly 3.0 m away from you makes a noise of 40.0 dB, what is the noise level of 1000 flies at that distance, assuming interference has a negligible effect?

Solution:

70.0 dB

Exercise:

Problem:

Ten cars in a circle at a boom box competition produce a 120-dB sound intensity level at the center of the circle. What is the average sound intensity level produced there by each stereo, assuming interference effects can be neglected?

Exercise:

Problem:

The amplitude of a sound wave is measured in terms of its maximum gauge pressure. By what factor does the amplitude of a sound wave increase if the sound intensity level goes up by 40.0 dB?

Solution:

100

Exercise:

Problem:

If a sound intensity level of 0 dB at 1000 Hz corresponds to a maximum gauge pressure (sound amplitude) of $10^{-9}\ \text{atm}$, what is the maximum gauge pressure in a 60-dB sound? What is the maximum gauge pressure in a 120-dB sound?

Exercise:

Problem:

An 8-hour exposure to a sound intensity level of 90.0 dB may cause hearing damage. What energy in joules falls on a 0.800-cm-diameter eardrum so exposed?

Solution:
Equation:

$$1.45 \times 10^{-3}\ \text{J}$$

Exercise:

Problem:

(a) Ear trumpets were never very common, but they did aid people with hearing losses by gathering sound over a large area and concentrating it on the smaller area of the eardrum. What decibel increase does an ear trumpet produce if its sound gathering area is $900\ \text{cm}^2$ and the area of the eardrum is $0.500\ \text{cm}^2$, but the trumpet only has an efficiency of 5.00% in transmitting the sound to the eardrum? (b) Comment on the usefulness of the decibel increase found in part (a).

Exercise:

Problem:

Sound is more effectively transmitted into a stethoscope by direct contact than through the air, and it is further intensified by being concentrated on the smaller area of the eardrum. It is reasonable to assume that sound is transmitted into a stethoscope 100 times as effectively compared with transmission though the air. What, then, is the gain in decibels produced by a stethoscope that has a sound gathering area of 15.0 cm^2, and concentrates the sound onto two eardrums with a total area of 0.900 cm^2 with an efficiency of 40.0%?

Solution:

28.2 dB

Exercise:

Problem:

Loudspeakers can produce intense sounds with surprisingly small energy input in spite of their low efficiencies. Calculate the power input needed to produce a 90.0-dB sound intensity level for a 12.0-cm-diameter speaker that has an efficiency of 1.00%. (This value is the sound intensity level right at the speaker.)

Glossary

intensity
: the power per unit area carried by a wave

sound intensity level
: a unitless quantity telling you the level of the sound relative to a fixed standard

sound pressure level
: the ratio of the pressure amplitude to a reference pressure

Doppler Effect and Sonic Booms

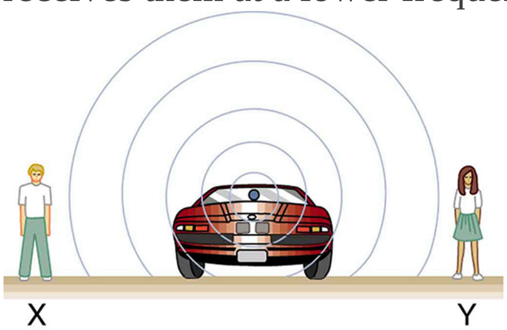
- Define Doppler effect, Doppler shift, and sonic boom.
- Calculate the frequency of a sound heard by someone observing Doppler shift.
- Describe the sounds produced by objects moving faster than the speed of sound.

The characteristic sound of a motorcycle buzzing by is an example of the **Doppler effect**. The high-pitch scream shifts dramatically to a lower-pitch roar as the motorcycle passes by a stationary observer. The closer the motorcycle brushes by, the more abrupt the shift. The faster the motorcycle moves, the greater the shift. We also hear this characteristic shift in frequency for passing race cars, airplanes, and trains. It is so familiar that it is used to imply motion and children often mimic it in play.

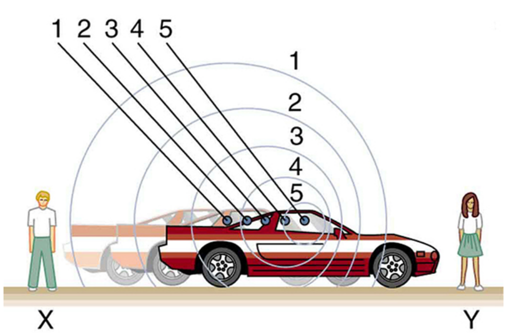
The Doppler effect is an alteration in the observed frequency of a sound due to motion of either the source or the observer. Although less familiar, this effect is easily noticed for a stationary source and moving observer. For example, if you ride a train past a stationary warning bell, you will hear the bell's frequency shift from high to low as you pass by. The actual change in frequency due to relative motion of source and observer is called a **Doppler shift**. The Doppler effect and Doppler shift are named for the Austrian physicist and mathematician Christian Johann Doppler (1803–1853), who did experiments with both moving sources and moving observers. Doppler, for example, had musicians play on a moving open train car and also play standing next to the train tracks as a train passed by. Their music was observed both on and off the train, and changes in frequency were measured.

What causes the Doppler shift? [link], [link], and [link] compare sound waves emitted by stationary and moving sources in a stationary air mass. Each disturbance spreads out spherically from the point where the sound was emitted. If the source is stationary, then all of the spheres representing the air compressions in the sound wave centered on the same point, and the stationary observers on either side see the same wavelength and frequency as emitted by the source, as in [link]. If the source is moving, as in [link], then the situation is different. Each compression of the air moves out in a

sphere from the point where it was emitted, but the point of emission moves. This moving emission point causes the air compressions to be closer together on one side and farther apart on the other. Thus, the wavelength is shorter in the direction the source is moving (on the right in [link]), and longer in the opposite direction (on the left in [link]). Finally, if the observers move, as in [link], the frequency at which they receive the compressions changes. The observer moving toward the source receives them at a higher frequency, and the person moving away from the source receives them at a lower frequency.

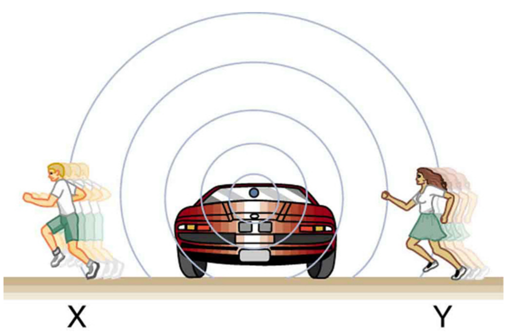


Sounds emitted by a source spread out in spherical waves. Because the source, observers, and air are stationary, the wavelength and frequency are the same in all directions and to all observers.



Sounds emitted by a

source moving to the right spread out from the points at which they were emitted. The wavelength is reduced and, consequently, the frequency is increased in the direction of motion, so that the observer on the right hears a higher-pitch sound. The opposite is true for the observer on the left, where the wavelength is increased and the frequency is reduced.



The same effect is produced when the observers move relative to the source. Motion toward the source increases frequency as the observer on the right passes through more wave crests than she would if stationary. Motion away from the

source decreases frequency as the observer on the left passes through fewer wave crests than he would if stationary.

We know that wavelength and frequency are related by $v_{\text{w}} = f\lambda$, where v_{w} is the fixed speed of sound. The sound moves in a medium and has the same speed v_{w} in that medium whether the source is moving or not. Thus f multiplied by λ is a constant. Because the observer on the right in [link] receives a shorter wavelength, the frequency she receives must be higher. Similarly, the observer on the left receives a longer wavelength, and hence he hears a lower frequency. The same thing happens in [link]. A higher frequency is received by the observer moving toward the source, and a lower frequency is received by an observer moving away from the source. In general, then, relative motion of source and observer toward one another increases the received frequency. Relative motion apart decreases frequency. The greater the relative speed is, the greater the effect.

Note:
The Doppler Effect
The Doppler effect occurs not only for sound but for any wave when there is relative motion between the observer and the source. There are Doppler shifts in the frequency of sound, light, and water waves, for example. Doppler shifts can be used to determine velocity, such as when ultrasound is reflected from blood in a medical diagnostic. The recession of galaxies is determined by the shift in the frequencies of light received from them and has implied much about the origins of the universe. Modern physics has been profoundly affected by observations of Doppler shifts.

For a stationary observer and a moving source, the frequency f_{obs} received by the observer can be shown to be

Equation:

$$f_{\text{obs}} = f_{\text{s}}\left(\frac{v_{\text{w}}}{v_{\text{w}} \pm v_{\text{s}}}\right),$$

where f_{s} is the frequency of the source, v_{s} is the speed of the source along a line joining the source and observer, and v_{w} is the speed of sound. The minus sign is used for motion toward the observer and the plus sign for motion away from the observer, producing the appropriate shifts up and down in frequency. Note that the greater the speed of the source, the greater the effect. Similarly, for a stationary source and moving observer, the frequency received by the observer f_{obs} is given by

Equation:

$$f_{\text{obs}} = f_{\text{s}}\left(\frac{v_{\text{w}} \pm v_{\text{obs}}}{v_{\text{w}}}\right),$$

where v_{obs} is the speed of the observer along a line joining the source and observer. Here the plus sign is for motion toward the source, and the minus is for motion away from the source.

Example:

Calculate Doppler Shift: A Train Horn

Suppose a train that has a 150-Hz horn is moving at 35.0 m/s in still air on a day when the speed of sound is 340 m/s.

(a) What frequencies are observed by a stationary person at the side of the tracks as the train approaches and after it passes?

(b) What frequency is observed by the train’s engineer traveling on the train?

Strategy

To find the observed frequency in (a), $f_{\text{obs}} = f_{\text{s}}\left(\frac{v_{\text{w}}}{v_{\text{w}} \pm v_{\text{s}}}\right)$, must be used because the source is moving. The minus sign is used for the approaching

train, and the plus sign for the receding train. In (b), there are two Doppler shifts—one for a moving source and the other for a moving observer.

Solution for (a)

(1) Enter known values into $f_{\text{obs}} = f_{\text{s}}\left(\frac{v_{\text{w}}}{v_{\text{w}} - v_{\text{s}}}\right)$.

Equation:

$$f_{\text{obs}} = f_{\text{s}}\left(\frac{v_{\text{w}}}{v_{\text{w}} - v_{\text{s}}}\right) = (150\ \text{Hz})\left(\frac{340\ \text{m/s}}{340\ \text{m/s} - 35.0\ \text{m/s}}\right)$$

(2) Calculate the frequency observed by a stationary person as the train approaches.

Equation:

$$f_{\text{obs}} = (150\ \text{Hz})(1.11) = 167\ \text{Hz}$$

(3) Use the same equation with the plus sign to find the frequency heard by a stationary person as the train recedes.

Equation:

$$f_{\text{obs}} = f_{\text{s}}\left(\frac{v_{\text{w}}}{v_{\text{w}} + v_{\text{s}}}\right) = (150\ \text{Hz})\left(\frac{340\ \text{m/s}}{340\ \text{m/s} + 35.0\ \text{m/s}}\right)$$

(4) Calculate the second frequency.

Equation:

$$f_{\text{obs}} = (150\ \text{Hz})(0.907) = 136\ \text{Hz}$$

Discussion on (a)

The numbers calculated are valid when the train is far enough away that the motion is nearly along the line joining train and observer. In both cases, the shift is significant and easily noticed. Note that the shift is 17.0 Hz for motion toward and 14.0 Hz for motion away. The shifts are not symmetric.

Solution for (b)

(1) Identify knowns:

- It seems reasonable that the engineer would receive the same frequency as emitted by the horn, because the relative velocity

between them is zero.
- Relative to the medium (air), the speeds are $v_s = v_{obs} = 35.0 \text{ m/s}$.
- The first Doppler shift is for the moving observer; the second is for the moving source.

(2) Use the following equation:
Equation:

$$f_{obs} = \left[f_s \left(\frac{v_w \pm v_{obs}}{v_w} \right) \right] \left(\frac{v_w}{v_w \pm v_s} \right).$$

The quantity in the square brackets is the Doppler-shifted frequency due to a moving observer. The factor on the right is the effect of the moving source.
(3) Because the train engineer is moving in the direction toward the horn, we must use the plus sign for v_{obs}; however, because the horn is also moving in the direction away from the engineer, we also use the plus sign for v_s. But the train is carrying both the engineer and the horn at the same velocity, so $v_s = v_{obs}$. As a result, everything but f_s cancels, yielding
Equation:

$$f_{obs}{=}f_s.$$

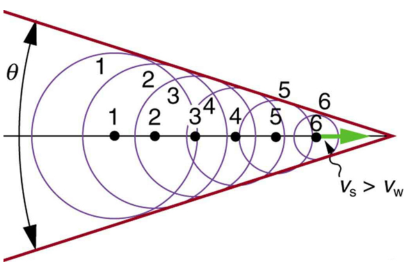
Discussion for (b)
We may expect that there is no change in frequency when source and observer move together because it fits your experience. For example, there is no Doppler shift in the frequency of conversations between driver and passenger on a motorcycle. People talking when a wind moves the air between them also observe no Doppler shift in their conversation. The crucial point is that source and observer are not moving relative to each other.

Sonic Booms to Bow Wakes

What happens to the sound produced by a moving source, such as a jet airplane, that approaches or even exceeds the speed of sound? The answer

to this question applies not only to sound but to all other waves as well.

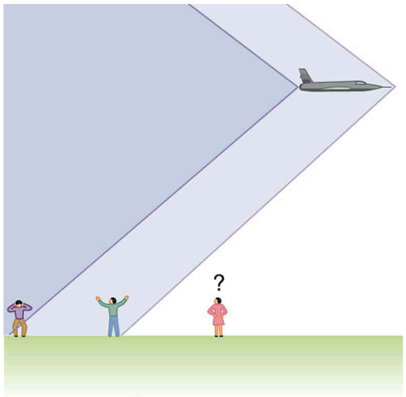
Suppose a jet airplane is coming nearly straight at you, emitting a sound of frequency f_s. The greater the plane's speed v_s, the greater the Doppler shift and the greater the value observed for f_{obs}. Now, as v_s approaches the speed of sound, f_{obs} approaches infinity, because the denominator in $f_{obs} = f_s\left(\frac{v_w}{v_w \pm v_s}\right)$ approaches zero. At the speed of sound, this result means that in front of the source, each successive wave is superimposed on the previous one because the source moves forward at the speed of sound. The observer gets them all at the same instant, and so the frequency is infinite. (Before airplanes exceeded the speed of sound, some people argued it would be impossible because such constructive superposition would produce pressures great enough to destroy the airplane.) If the source exceeds the speed of sound, no sound is received by the observer until the source has passed, so that the sounds from the approaching source are mixed with those from it when receding. This mixing appears messy, but something interesting happens—a sonic boom is created. (See [link].)



Sound waves from a source that moves faster than the speed of sound spread spherically from the point where they are emitted, but the source moves ahead of each.

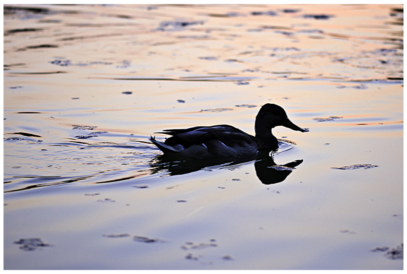
Constructive interference along the lines shown (actually a cone in three dimensions) creates a shock wave called a sonic boom. The faster the speed of the source, the smaller the angle θ.

There is constructive interference along the lines shown (a cone in three dimensions) from similar sound waves arriving there simultaneously. This superposition forms a disturbance called a **sonic boom**, a constructive interference of sound created by an object moving faster than sound. Inside the cone, the interference is mostly destructive, and so the sound intensity there is much less than on the shock wave. An aircraft creates two sonic booms, one from its nose and one from its tail. (See [link].) During television coverage of space shuttle landings, two distinct booms could often be heard. These were separated by exactly the time it would take the shuttle to pass by a point. Observers on the ground often do not see the aircraft creating the sonic boom, because it has passed by before the shock wave reaches them, as seen in [link]. If the aircraft flies close by at low altitude, pressures in the sonic boom can be destructive and break windows as well as rattle nerves. Because of how destructive sonic booms can be, supersonic flights are banned over populated areas of the United States.

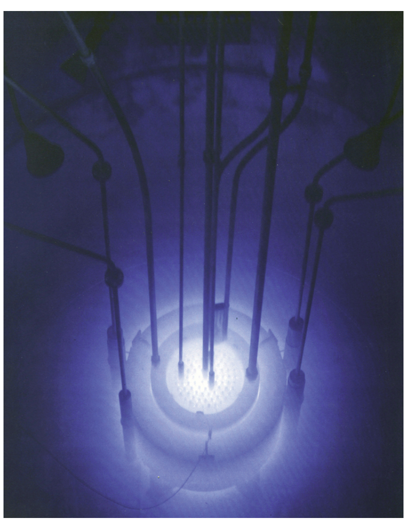


Two sonic booms, created by the nose and tail of an aircraft, are observed on the ground after the plane has passed by.

Sonic booms are one example of a broader phenomenon called bow wakes. A **bow wake**, such as the one in [link], is created when the wave source moves faster than the wave propagation speed. Water waves spread out in circles from the point where created, and the bow wake is the familiar V-shaped wake trailing the source. A more exotic bow wake is created when a subatomic particle travels through a medium faster than the speed of light travels in that medium. (In a vacuum, the maximum speed of light will be $c = 3.00 \times 10^8 \text{ m/s}$; in the medium of water, the speed of light is closer to $0.75c$. If the particle creates light in its passage, that light spreads on a cone with an angle indicative of the speed of the particle, as illustrated in [link]. Such a bow wake is called Cerenkov radiation and is commonly observed in particle physics.



Bow wake created by a duck. Constructive interference produces the rather structured wake, while there is relatively little wave action inside the wake, where interference is mostly destructive. (credit: Horia Varlan, Flickr)



The blue glow in this research reactor pool is Cerenkov radiation caused by subatomic particles traveling faster than the speed of light in water. (credit: U.S. Nuclear Regulatory Commission)

Doppler shifts and sonic booms are interesting sound phenomena that occur in all types of waves. They can be of considerable use. For example, the Doppler shift in ultrasound can be used to measure blood velocity, while police use the Doppler shift in radar (a microwave) to measure car velocities. In meteorology, the Doppler shift is used to track the motion of storm clouds; such “Doppler Radar” can give velocity and direction and rain or snow potential of imposing weather fronts. In astronomy, we can examine the light emitted from distant galaxies and determine their speed relative to ours. As galaxies move away from us, their light is shifted to a lower frequency, and so to a longer wavelength—the so-called red shift. Such information from galaxies far, far away has allowed us to estimate the age of the universe (from the Big Bang) as about 14 billion years.

Exercise:

Check Your Understanding

Problem:

Why did scientist Christian Doppler observe musicians both on a moving train and also from a stationary point not on the train?

Solution:

Doppler needed to compare the perception of sound when the observer is stationary and the sound source moves, as well as when the sound

source and the observer are both in motion.

Exercise:
Check Your Understanding

Problem:

Describe a situation in your life when you might rely on the Doppler shift to help you either while driving a car or walking near traffic.

Solution:

If I am driving and I hear Doppler shift in an ambulance siren, I would be able to tell when it was getting closer and also if it has passed by. This would help me to know whether I needed to pull over and let the ambulance through.

Section Summary

- The Doppler effect is an alteration in the observed frequency of a sound due to motion of either the source or the observer.
- The actual change in frequency is called the Doppler shift.
- A sonic boom is constructive interference of sound created by an object moving faster than sound.
- A sonic boom is a type of bow wake created when any wave source moves faster than the wave propagation speed.
- For a stationary observer and a moving source, the observed frequency f_{obs} is:
  Equation:
  $$f_{\text{obs}} = f_{\text{s}} \left(\frac{v_{\text{w}}}{v_{\text{w}} \pm v_{\text{s}}} \right),$$
  where f_{s} is the frequency of the source, v_{s} is the speed of the source, and v_{w} is the speed of sound. The minus sign is used for motion toward the observer and the plus sign for motion away.
- For a stationary source and moving observer, the observed frequency is:

Equation:

$$f_{\text{obs}} = f_{\text{s}}\left(\frac{v_{\text{w}} \pm v_{\text{obs}}}{v_{\text{w}}}\right),$$

where v_{obs} is the speed of the observer.

Conceptual Questions

Exercise:

Problem: Is the Doppler shift real or just a sensory illusion?

Exercise:

Problem:

Due to efficiency considerations related to its bow wake, the supersonic transport aircraft must maintain a cruising speed that is a constant ratio to the speed of sound (a constant Mach number). If the aircraft flies from warm air into colder air, should it increase or decrease its speed? Explain your answer.

Exercise:

Problem:

When you hear a sonic boom, you often cannot see the plane that made it. Why is that?

Problems & Exercises

Exercise:

Problem:

(a) What frequency is received by a person watching an oncoming ambulance moving at 110 km/h and emitting a steady 800-Hz sound from its siren? The speed of sound on this day is 345 m/s. (b) What frequency does she receive after the ambulance has passed?

Solution:

(a) 878 Hz

(b) 735 Hz

Exercise:

Problem:

(a) At an air show a jet flies directly toward the stands at a speed of 1200 km/h, emitting a frequency of 3500 Hz, on a day when the speed of sound is 342 m/s. What frequency is received by the observers? (b) What frequency do they receive as the plane flies directly away from them?

Exercise:

Problem:

What frequency is received by a mouse just before being dispatched by a hawk flying at it at 25.0 m/s and emitting a screech of frequency 3500 Hz? Take the speed of sound to be 331 m/s.

Solution:
Equation:

$$3.79 \times 10^3 \text{ Hz}$$

Exercise:

Problem:

A spectator at a parade receives an 888-Hz tone from an oncoming trumpeter who is playing an 880-Hz note. At what speed is the musician approaching if the speed of sound is 338 m/s?

Exercise:

Problem:

A commuter train blows its 200-Hz horn as it approaches a crossing. The speed of sound is 335 m/s. (a) An observer waiting at the crossing receives a frequency of 208 Hz. What is the speed of the train? (b) What frequency does the observer receive as the train moves away?

Solution:

(a) 12.9 m/s

(b) 193 Hz

Exercise:

Problem:

Can you perceive the shift in frequency produced when you pull a tuning fork toward you at 10.0 m/s on a day when the speed of sound is 344 m/s? To answer this question, calculate the factor by which the frequency shifts and see if it is greater than 0.300%.

Exercise:

Problem:

Two eagles fly directly toward one another, the first at 15.0 m/s and the second at 20.0 m/s. Both screech, the first one emitting a frequency of 3200 Hz and the second one emitting a frequency of 3800 Hz. What frequencies do they receive if the speed of sound is 330 m/s?

Solution:

First eagle hears 4.23×10^3 Hz

Second eagle hears 3.56×10^3 Hz

Exercise:

Problem:

What is the minimum speed at which a source must travel toward you for you to be able to hear that its frequency is Doppler shifted? That is, what speed produces a shift of 0.300% on a day when the speed of sound is 331 m/s?

Glossary

Doppler effect
: an alteration in the observed frequency of a sound due to motion of either the source or the observer

Doppler shift
: the actual change in frequency due to relative motion of source and observer

sonic boom
: a constructive interference of sound created by an object moving faster than sound

bow wake
: V-shaped disturbance created when the wave source moves faster than the wave propagation speed

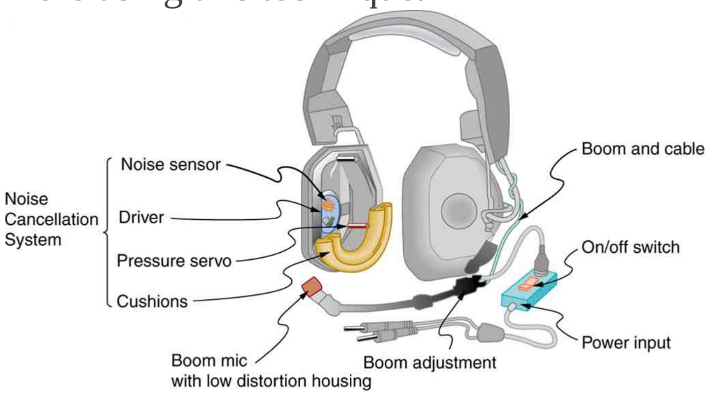
Sound Interference and Resonance: Standing Waves in Air Columns

- Define antinode, node, fundamental, overtones, and harmonics.
- Identify instances of sound interference in everyday situations.
- Describe how sound interference occurring inside open and closed tubes changes the characteristics of the sound, and how this applies to sounds produced by musical instruments.
- Calculate the length of a tube using sound wave measurements.

Some types of headphones use the phenomena of constructive and destructive interference to cancel out outside noises. (credit: JVC America, Flickr)

Interference is the hallmark of waves, all of which exhibit constructive and destructive interference exactly analogous to that seen for water waves. In fact, one way to prove something “is a wave” is to observe interference effects. So, sound being a wave, we expect it to exhibit interference; we have already mentioned a few such effects, such as the beats from two similar notes played simultaneously.

[link] shows a clever use of sound interference to cancel noise. Larger-scale applications of active noise reduction by destructive interference are contemplated for entire passenger compartments in commercial aircraft. To obtain destructive interference, a fast electronic analysis is performed, and a second sound is introduced with its maxima and minima exactly reversed from the incoming noise. Sound waves in fluids are pressure waves and consistent with Pascal’s principle; pressures from two different sources add and subtract like simple numbers; that is, positive and negative gauge pressures add to a much smaller pressure, producing a lower-intensity sound. Although completely destructive interference is possible only under the simplest conditions, it is possible to reduce noise levels by 30 dB or more using this technique.



Headphones designed to cancel noise with destructive interference create a sound wave exactly opposite to the incoming sound. These headphones can be more effective than the simple passive attenuation used in most ear protection. Such headphones were

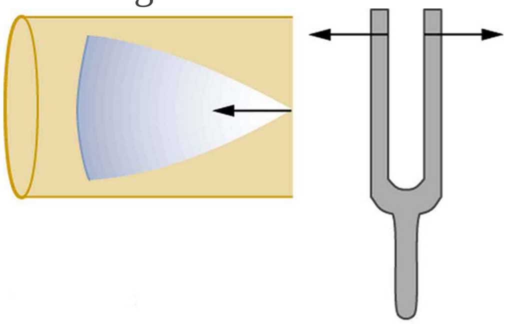
used on the record-setting, around the world nonstop flight of the Voyager aircraft to protect the pilots' hearing from engine noise.

Where else can we observe sound interference? All sound resonances, such as in musical instruments, are due to constructive and destructive interference. Only the resonant frequencies interfere constructively to form standing waves, while others interfere destructively and are absent. From the toot made by blowing over a bottle, to the characteristic flavor of a violin's sounding box, to the recognizability of a great singer's voice, resonance and standing waves play a vital role.

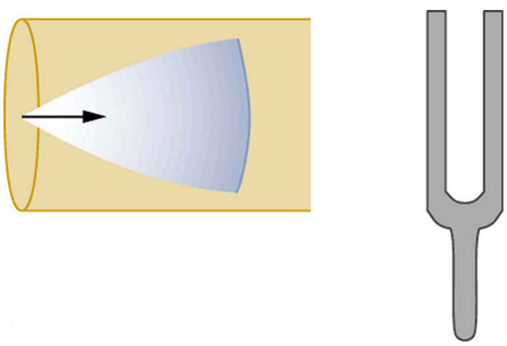
Note:
Interference
Interference is such a fundamental aspect of waves that observing interference is proof that something is a wave. The wave nature of light was established by experiments showing interference. Similarly, when electrons scattered from crystals exhibited interference, their wave nature was confirmed to be exactly as predicted by symmetry with certain wave characteristics of light.

Suppose we hold a tuning fork near the end of a tube that is closed at the other end, as shown in [link], [link], [link], and [link]. If the tuning fork has just the right frequency, the air column in the tube resonates loudly, but at most frequencies it vibrates very little. This observation just means that the air column has only certain natural frequencies. The figures show how a resonance at the lowest of these natural frequencies is formed. A disturbance travels down the tube at the speed of sound and bounces off the closed end. If the tube is just the right length, the reflected sound arrives back at the tuning fork exactly half a cycle later, and it interferes

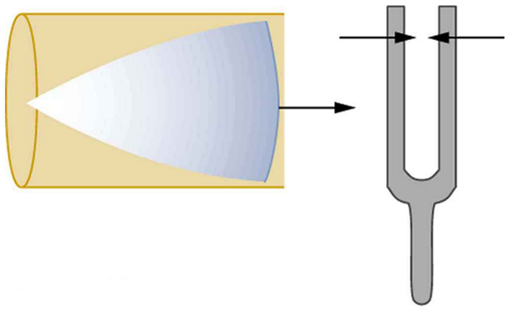
constructively with the continuing sound produced by the tuning fork. The incoming and reflected sounds form a standing wave in the tube as shown.



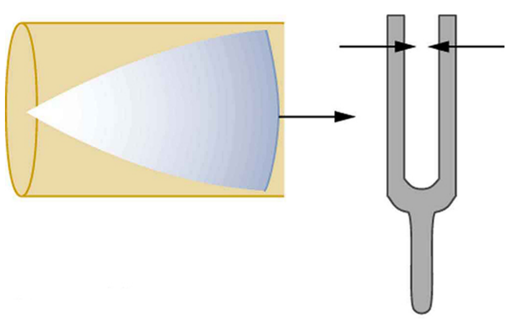
Resonance of air in a tube closed at one end, caused by a tuning fork. A disturbance moves down the tube.



Resonance of air in a tube closed at one end, caused by a tuning fork. The disturbance reflects from the closed end of the tube.



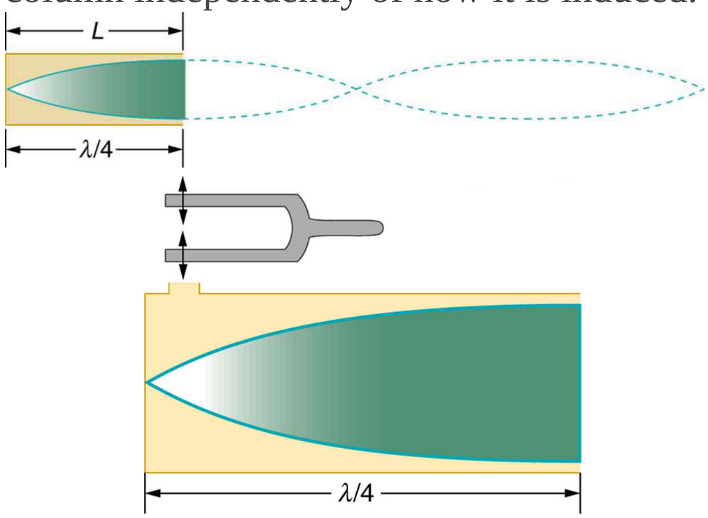
Resonance of air in a tube closed at one end, caused by a tuning fork. If the length of the tube L is just right, the disturbance gets back to the tuning fork half a cycle later and interferes constructively with the continuing sound from the tuning fork. This interference forms a standing wave, and the air column resonates.



Resonance of air in a tube closed at one end, caused by a tuning fork. A graph of air displacement along the length of the tube shows none at the closed

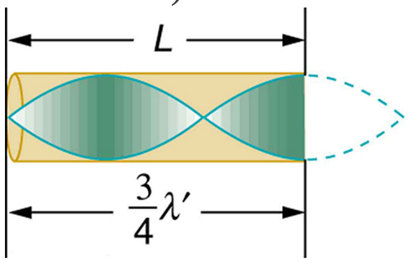
end, where the motion is constrained, and a maximum at the open end. This standing wave has one-fourth of its wavelength in the tube, so that $\lambda = 4L$.

The standing wave formed in the tube has its maximum air displacement (an **antinode**) at the open end, where motion is unconstrained, and no displacement (a **node**) at the closed end, where air movement is halted. The distance from a node to an antinode is one-fourth of a wavelength, and this equals the length of the tube; thus, $\lambda = 4L$. This same resonance can be produced by a vibration introduced at or near the closed end of the tube, as shown in [link]. It is best to consider this a natural vibration of the air column independently of how it is induced.

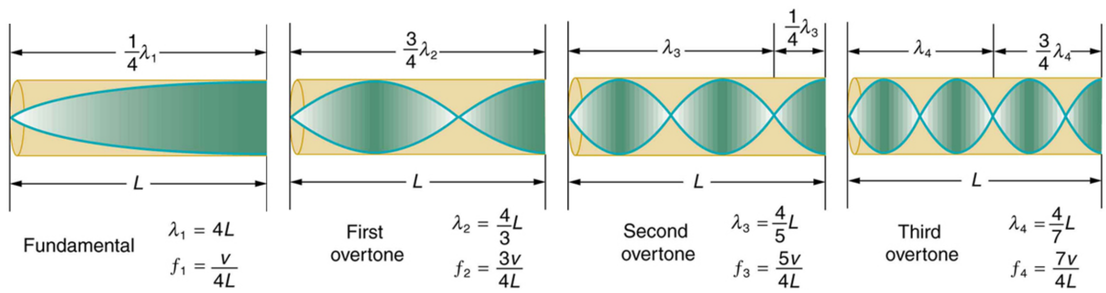


The same standing wave is created in the tube by a vibration introduced near its closed end.

Given that maximum air displacements are possible at the open end and none at the closed end, there are other, shorter wavelengths that can resonate in the tube, such as the one shown in [link]. Here the standing wave has three-fourths of its wavelength in the tube, or $L = (3/4)\lambda\prime$, so that $\lambda\prime = 4L/3$. Continuing this process reveals a whole series of shorter-wavelength and higher-frequency sounds that resonate in the tube. We use specific terms for the resonances in any system. The lowest resonant frequency is called the **fundamental**, while all higher resonant frequencies are called **overtones**. All resonant frequencies are integral multiples of the fundamental, and they are collectively called **harmonics**. The fundamental is the first harmonic, the first overtone is the second harmonic, and so on. [link] shows the fundamental and the first three overtones (the first four harmonics) in a tube closed at one end.

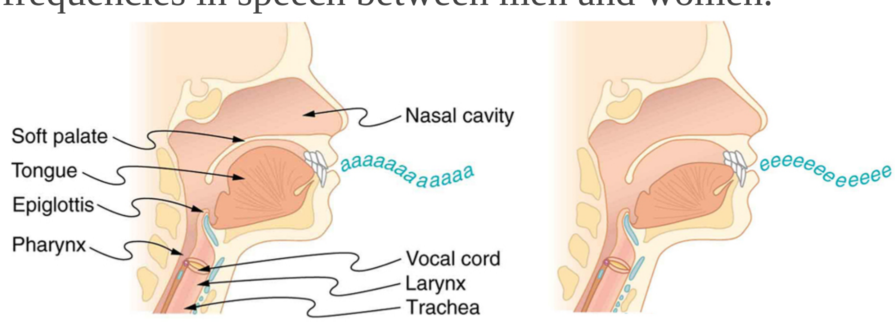


Another resonance for a tube closed at one end. This has maximum air displacements at the open end, and none at the closed end. The wavelength is shorter, with three-fourths $\lambda\prime$ equaling the length of the tube, so that $\lambda\prime = 4L/3$. This higher-frequency vibration is the first overtone.



The fundamental and three lowest overtones for a tube closed at one end. All have maximum air displacements at the open end and none at the closed end.

The fundamental and overtones can be present simultaneously in a variety of combinations. For example, middle C on a trumpet has a sound distinctively different from middle C on a clarinet, both instruments being modified versions of a tube closed at one end. The fundamental frequency is the same (and usually the most intense), but the overtones and their mix of intensities are different and subject to shading by the musician. This mix is what gives various musical instruments (and human voices) their distinctive characteristics, whether they have air columns, strings, sounding boxes, or drumheads. In fact, much of our speech is determined by shaping the cavity formed by the throat and mouth and positioning the tongue to adjust the fundamental and combination of overtones. Simple resonant cavities can be made to resonate with the sound of the vowels, for example. (See [link].) In boys, at puberty, the larynx grows and the shape of the resonant cavity changes giving rise to the difference in predominant frequencies in speech between men and women.



The throat and mouth form an air column closed at one end that resonates in response to vibrations in the voice box. The spectrum of overtones and their intensities vary with mouth shaping and tongue position to form different sounds. The voice box can be replaced with a mechanical vibrator, and understandable speech is still possible. Variations in basic shapes make different voices recognizable.

Now let us look for a pattern in the resonant frequencies for a simple tube that is closed at one end. The fundamental has $\lambda = 4L$, and frequency is related to wavelength and the speed of sound as given by:

Equation:

$$v_{\text{w}} = f\lambda.$$

Solving for f in this equation gives

Equation:

$$f = \frac{v_{\text{w}}}{\lambda} = \frac{v_{\text{w}}}{4L},$$

where v_{w} is the speed of sound in air. Similarly, the first overtone has $\lambda\prime = 4L/3$ (see [link]), so that

Equation:

$$f\prime = 3\frac{v_{\text{w}}}{4L} = 3f.$$

Because $f\prime = 3f$, we call the first overtone the third harmonic. Continuing this process, we see a pattern that can be generalized in a single expression. The resonant frequencies of a tube closed at one end are

Equation:

$$f_n = n\frac{v_w}{4L}, n = 1,3,5,$$

where f_1 is the fundamental, f_3 is the first overtone, and so on. It is interesting that the resonant frequencies depend on the speed of sound and, hence, on temperature. This dependence poses a noticeable problem for organs in old unheated cathedrals, and it is also the reason why musicians commonly bring their wind instruments to room temperature before playing them.

Example:
Find the Length of a Tube with a 128 Hz Fundamental
(a) What length should a tube closed at one end have on a day when the air temperature, is $22.0^{\circ}C$, if its fundamental frequency is to be 128 Hz (C below middle C)?
(b) What is the frequency of its fourth overtone?
Strategy
The length L can be found from the relationship in $f_n = n\frac{v_w}{4L}$, but we will first need to find the speed of sound v_w.
Solution for (a)
(1) Identify knowns:

- the fundamental frequency is 128 Hz
- the air temperature is $22.0^{\circ}C$

(2) Use $f_n = n\frac{v_w}{4L}$ to find the fundamental frequency ($n = 1$).
Equation:

$$f_1 = \frac{v_w}{4L}$$

(3) Solve this equation for length.
Equation:

$$L = \frac{v_w}{4f_1}$$

(4) Find the speed of sound using $v_{\mathrm{w}} = (331\ \mathrm{m/s})\sqrt{\frac{T}{273\ \mathrm{K}}}$.

Equation:

$$v_{\mathrm{w}} = (331\ \mathrm{m/s})\sqrt{\frac{295\ \mathrm{K}}{273\ \mathrm{K}}} = 344\ \mathrm{m/s}$$

(5) Enter the values of the speed of sound and frequency into the expression for L.

Equation:

$$L = \frac{v_{\mathrm{w}}}{4f_1} = \frac{344\ \mathrm{m/s}}{4(128\ \mathrm{Hz})} = 0.672\ \mathrm{m}$$

Discussion on (a)

Many wind instruments are modified tubes that have finger holes, valves, and other devices for changing the length of the resonating air column and hence, the frequency of the note played. Horns producing very low frequencies, such as tubas, require tubes so long that they are coiled into loops.

Solution for (b)

(1) Identify knowns:

- the first overtone has $n = 3$
- the second overtone has $n = 5$
- the third overtone has $n = 7$
- the fourth overtone has $n = 9$

(2) Enter the value for the fourth overtone into $f_n = n\frac{v_{\mathrm{w}}}{4L}$.

Equation:

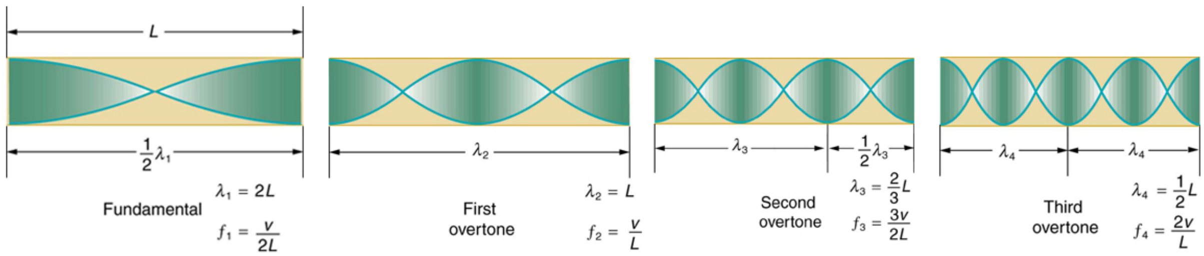
$$f_9 = 9\frac{v_{\mathrm{w}}}{4L} = 9f_1 = 1.15\ \mathrm{kHz}$$

Discussion on (b)

Whether this overtone occurs in a simple tube or a musical instrument depends on how it is stimulated to vibrate and the details of its shape. The

trombone, for example, does not produce its fundamental frequency and only makes overtones.

Another type of tube is one that is *open* at both ends. Examples are some organ pipes, flutes, and oboes. The resonances of tubes open at both ends can be analyzed in a very similar fashion to those for tubes closed at one end. The air columns in tubes open at both ends have maximum air displacements at both ends, as illustrated in [link]. Standing waves form as shown.



The resonant frequencies of a tube open at both ends are shown, including the fundamental and the first three overtones. In all cases the maximum air displacements occur at both ends of the tube, giving it different natural frequencies than a tube closed at one end.

Based on the fact that a tube open at both ends has maximum air displacements at both ends, and using [link] as a guide, we can see that the resonant frequencies of a tube open at both ends are:

Equation:

$$f_n = n\frac{v_{\text{w}}}{2L},\ n = 1, 2, 3...,$$

where f_1 is the fundamental, f_2 is the first overtone, f_3 is the second overtone, and so on. Note that a tube open at both ends has a fundamental frequency twice what it would have if closed at one end. It also has a different spectrum of overtones than a tube closed at one end. So if you had

two tubes with the same fundamental frequency but one was open at both ends and the other was closed at one end, they would sound different when played because they have different overtones. Middle C, for example, would sound richer played on an open tube, because it has even multiples of the fundamental as well as odd. A closed tube has only odd multiples.

Note:
Real-World Applications: Resonance in Everyday Systems
Resonance occurs in many different systems, including strings, air columns, and atoms. Resonance is the driven or forced oscillation of a system at its natural frequency. At resonance, energy is transferred rapidly to the oscillating system, and the amplitude of its oscillations grows until the system can no longer be described by Hooke's law. An example of this is the distorted sound intentionally produced in certain types of rock music.

Wind instruments use resonance in air columns to amplify tones made by lips or vibrating reeds. Other instruments also use air resonance in clever ways to amplify sound. [link] shows a violin and a guitar, both of which have sounding boxes but with different shapes, resulting in different overtone structures. The vibrating string creates a sound that resonates in the sounding box, greatly amplifying the sound and creating overtones that give the instrument its characteristic flavor. The more complex the shape of the sounding box, the greater its ability to resonate over a wide range of frequencies. The marimba, like the one shown in [link] uses pots or gourds below the wooden slats to amplify their tones. The resonance of the pot can be adjusted by adding water.

String instruments such as violins and guitars use resonance in their sounding boxes to amplify and enrich the sound created by their vibrating strings. The bridge and supports couple the string vibrations to the sounding boxes and air within. (credits: guitar, Feliciano Guimares, Fotopedia; violin, Steve Snodgrass, Flickr)

Resonance has been used in musical instruments since prehistoric times. This marimba uses gourds as resonance chambers to amplify its sound. (credit: APC Events, Flickr)

We have emphasized sound applications in our discussions of resonance and standing waves, but these ideas apply to any system that has wave characteristics. Vibrating strings, for example, are actually resonating and have fundamentals and overtones similar to those for air columns. More subtle are the resonances in atoms due to the wave character of their electrons. Their orbitals can be viewed as standing waves, which have a fundamental (ground state) and overtones (excited states). It is fascinating that wave characteristics apply to such a wide range of physical systems.

Exercise:

Check Your Understanding

Problem:

Describe how noise-canceling headphones differ from standard headphones used to block outside sounds.

Solution:

Regular headphones only block sound waves with a physical barrier. Noise-canceling headphones use destructive interference to reduce the loudness of outside sounds.

Exercise:
Check Your Understanding

Problem:

How is it possible to use a standing wave's node and antinode to determine the length of a closed-end tube?

Solution:

When the tube resonates at its natural frequency, the wave's node is located at the closed end of the tube, and the antinode is located at the open end. The length of the tube is equal to one-fourth of the wavelength of this wave. Thus, if we know the wavelength of the wave, we can determine the length of the tube.

Note:
PhET Explorations: Sound
This simulation lets you see sound waves. Adjust the frequency or volume and you can see and hear how the wave changes. Move the listener around and hear what she hears.
https://archive.cnx.org/specials/c4d3b96e-41f3-11e5-ab7b-47e22dffc18e/sound/#sim-single-source

Section Summary

- Sound interference and resonance have the same properties as defined for all waves.
- In air columns, the lowest-frequency resonance is called the fundamental, whereas all higher resonant frequencies are called overtones. Collectively, they are called harmonics.

- The resonant frequencies of a tube closed at one end are:
  Equation:

$$f_n = n\frac{v_{\mathrm{w}}}{4L}, n = 1, 3, 5...,$$

  f_1 is the fundamental and L is the length of the tube.
- The resonant frequencies of a tube open at both ends are:
  Equation:

$$f_n = n\frac{v_{\mathrm{w}}}{2L}, n = 1, 2, 3...$$

Conceptual Questions

Exercise:

Problem:

How does an unamplified guitar produce sounds so much more intense than those of a plucked string held taut by a simple stick?

Exercise:

Problem:

You are given two wind instruments of identical length. One is open at both ends, whereas the other is closed at one end. Which is able to produce the lowest frequency?

Exercise:

Problem:

What is the difference between an overtone and a harmonic? Are all harmonics overtones? Are all overtones harmonics?

Problems & Exercises

Exercise:

Problem:

A “showy” custom-built car has two brass horns that are supposed to produce the same frequency but actually emit 263.8 and 264.5 Hz. What beat frequency is produced?

Solution:

0.7 Hz

Exercise:

Problem:

What beat frequencies will be present: (a) If the musical notes A and C are played together (frequencies of 220 and 264 Hz)? (b) If D and F are played together (frequencies of 297 and 352 Hz)? (c) If all four are played together?

Exercise:

Problem:

What beat frequencies result if a piano hammer hits three strings that emit frequencies of 127.8, 128.1, and 128.3 Hz?

Solution:

0.3 Hz, 0.2 Hz, 0.5 Hz

Exercise:

Problem:

A piano tuner hears a beat every 2.00 s when listening to a 264.0-Hz tuning fork and a single piano string. What are the two possible frequencies of the string?

Exercise:

Problem:

(a) What is the fundamental frequency of a 0.672-m-long tube, open at both ends, on a day when the speed of sound is 344 m/s? (b) What is the frequency of its second harmonic?

Solution:

(a) 256 Hz

(b) 512 Hz

Exercise:

Problem:

If a wind instrument, such as a tuba, has a fundamental frequency of 32.0 Hz, what are its first three overtones? It is closed at one end. (The overtones of a real tuba are more complex than this example, because it is a tapered tube.)

Exercise:

Problem:

What are the first three overtones of a bassoon that has a fundamental frequency of 90.0 Hz? It is open at both ends. (The overtones of a real bassoon are more complex than this example, because its double reed makes it act more like a tube closed at one end.)

Solution:

180 Hz, 270 Hz, 360 Hz

Exercise:

Problem:

How long must a flute be in order to have a fundamental frequency of 262 Hz (this frequency corresponds to middle C on the evenly tempered chromatic scale) on a day when air temperature is $20.0^{\circ}C$? It is open at both ends.

Exercise:

Problem:

What length should an oboe have to produce a fundamental frequency of 110 Hz on a day when the speed of sound is 343 m/s? It is open at both ends.

Solution:

1.56 m

Exercise:

Problem:

What is the length of a tube that has a fundamental frequency of 176 Hz and a first overtone of 352 Hz if the speed of sound is 343 m/s?

Exercise:

Problem:

(a) Find the length of an organ pipe closed at one end that produces a fundamental frequency of 256 Hz when air temperature is $18.0^{\circ}C$. (b) What is its fundamental frequency at $25.0^{\circ}C$?

Solution:

(a) 0.334 m

(b) 259 Hz

Exercise:

Problem:

By what fraction will the frequencies produced by a wind instrument change when air temperature goes from 10.0°C to 30.0°C? That is, find the ratio of the frequencies at those temperatures.

Exercise:

Problem:

The ear canal resonates like a tube closed at one end. (See [link].) If ear canals range in length from 1.80 to 2.60 cm in an average population, what is the range of fundamental resonant frequencies? Take air temperature to be 37.0°C, which is the same as body temperature. How does this result correlate with the intensity versus frequency graph ([link] of the human ear?

Solution:

3.39 to 4.90 kHz

Exercise:

Problem:

Calculate the first overtone in an ear canal, which resonates like a 2.40-cm-long tube closed at one end, by taking air temperature to be 37.0°C. Is the ear particularly sensitive to such a frequency? (The resonances of the ear canal are complicated by its nonuniform shape, which we shall ignore.)

Exercise:

Problem:

A crude approximation of voice production is to consider the breathing passages and mouth to be a resonating tube closed at one end. (See [link].) (a) What is the fundamental frequency if the tube is 0.240-m long, by taking air temperature to be 37.0°C? (b) What would this frequency become if the person replaced the air with helium? Assume the same temperature dependence for helium as for air.

Solution:

(a) 367 Hz

(b) 1.07 kHz

Exercise:

Problem:

(a) Students in a physics lab are asked to find the length of an air column in a tube closed at one end that has a fundamental frequency of 256 Hz. They hold the tube vertically and fill it with water to the top, then lower the water while a 256-Hz tuning fork is rung and listen for the first resonance. What is the air temperature if the resonance occurs for a length of 0.336 m? (b) At what length will they observe the second resonance (first overtone)?

Exercise:

Problem:

What frequencies will a 1.80-m-long tube produce in the audible range at 20.0°C if: (a) The tube is closed at one end? (b) It is open at both ends?

Solution:

(a) $f_n = n(47.6\text{ Hz}),\ n = 1,\ 3,\ 5, ...,\ 419$

(b) $f_n = n(95.3\text{ Hz}),\ n = 1,\ 2,\ 3, ...,\ 210$

Glossary

antinode
: point of maximum displacement

node

point of zero displacement

fundamental
the lowest-frequency resonance

overtones
all resonant frequencies higher than the fundamental

harmonics
the term used to refer collectively to the fundamental and its overtones

Hearing

- Define hearing, pitch, loudness, timbre, note, tone, phon, ultrasound, and infrasound.
- Compare loudness to frequency and intensity of a sound.
- Identify structures of the inner ear and explain how they relate to sound perception.

Hearing allows this vocalist, his band, and his fans to enjoy music. (credit: West Point Public Affairs, Flickr)

The human ear has a tremendous range and sensitivity. It can give us a wealth of simple information—such as pitch, loudness, and direction. And from its input we can detect musical quality and nuances of voiced emotion. How is our hearing related to the physical qualities of sound, and how does the hearing mechanism work?

Hearing is the perception of sound. (Perception is commonly defined to be awareness through the senses, a typically circular definition of higher-level processes in living organisms.) Normal human hearing encompasses frequencies from 20 to 20,000 Hz, an impressive range. Sounds below 20 Hz are called **infrasound**, whereas those above 20,000 Hz are **ultrasound**. Neither is perceived by the ear, although infrasound can sometimes be felt as vibrations. When we do hear low-frequency vibrations, such as the

sounds of a diving board, we hear the individual vibrations only because there are higher-frequency sounds in each. Other animals have hearing ranges different from that of humans. Dogs can hear sounds as high as 30,000 Hz, whereas bats and dolphins can hear up to 100,000-Hz sounds. You may have noticed that dogs respond to the sound of a dog whistle which produces sound out of the range of human hearing. Elephants are known to respond to frequencies below 20 Hz.

The perception of frequency is called **pitch**. Most of us have excellent relative pitch, which means that we can tell whether one sound has a different frequency from another. Typically, we can discriminate between two sounds if their frequencies differ by 0.3% or more. For example, 500.0 and 501.5 Hz are noticeably different. Pitch perception is directly related to frequency and is not greatly affected by other physical quantities such as intensity. Musical **notes** are particular sounds that can be produced by most instruments and in Western music have particular names. Combinations of notes constitute music. Some people can identify musical notes, such as A-sharp, C, or E-flat, just by listening to them. This uncommon ability is called perfect pitch.

The ear is remarkably sensitive to low-intensity sounds. The lowest audible intensity or threshold is about $10^{-12}\ \mathrm{W/m^2}$ or 0 dB. Sounds as much as 10^{12} more intense can be briefly tolerated. Very few measuring devices are capable of observations over a range of a trillion. The perception of intensity is called **loudness**. At a given frequency, it is possible to discern differences of about 1 dB, and a change of 3 dB is easily noticed. But loudness is not related to intensity alone. Frequency has a major effect on how loud a sound seems. The ear has its maximum sensitivity to frequencies in the range of 2000 to 5000 Hz, so that sounds in this range are perceived as being louder than, say, those at 500 or 10,000 Hz, even when they all have the same intensity. Sounds near the high- and low-frequency extremes of the hearing range seem even less loud, because the ear is even less sensitive at those frequencies. [link] gives the dependence of certain human hearing perceptions on physical quantities.

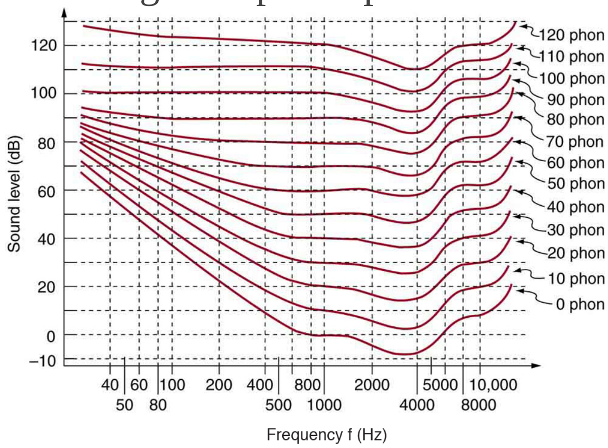
Perception	**Physical quantity**
Pitch	Frequency
Loudness	Intensity and Frequency
Timbre	Number and relative intensity of multiple frequencies. Subtle craftsmanship leads to non-linear effects and more detail.
Note	Basic unit of music with specific names, combined to generate tunes
Tone	Number and relative intensity of multiple frequencies.

Sound Perceptions

When a violin plays middle C, there is no mistaking it for a piano playing the same note. The reason is that each instrument produces a distinctive set of frequencies and intensities. We call our perception of these combinations of frequencies and intensities **tone** quality, or more commonly the **timbre** of the sound. It is more difficult to correlate timbre perception to physical quantities than it is for loudness or pitch perception. Timbre is more subjective. Terms such as dull, brilliant, warm, cold, pure, and rich are employed to describe the timbre of a sound. So the consideration of timbre takes us into the realm of perceptual psychology, where higher-level processes in the brain are dominant. This is true for other perceptions of sound, such as music and noise. We shall not delve further into them; rather, we will concentrate on the question of loudness perception.

A unit called a **phon** is used to express loudness numerically. Phons differ from decibels because the phon is a unit of loudness perception, whereas the decibel is a unit of physical intensity. [link] shows the relationship of loudness to intensity (or intensity level) and frequency for persons with normal hearing. The curved lines are equal-loudness curves. Each curve is

labeled with its loudness in phons. Any sound along a given curve will be perceived as equally loud by the average person. The curves were determined by having large numbers of people compare the loudness of sounds at different frequencies and sound intensity levels. At a frequency of 1000 Hz, phons are taken to be numerically equal to decibels. The following example helps illustrate how to use the graph:



The relationship of loudness in phons to intensity level (in decibels) and intensity (in watts per meter squared) for persons with normal hearing. The curved lines are equal-loudness curves—all sounds on a given curve are perceived as equally loud. Phons and decibels are defined to be the same at 1000 Hz.

Example:
Measuring Loudness: Loudness Versus Intensity Level and Frequency
(a) What is the loudness in phons of a 100-Hz sound that has an intensity level of 80 dB? (b) What is the intensity level in decibels of a 4000-Hz

sound having a loudness of 70 phons? (c) At what intensity level will an 8000-Hz sound have the same loudness as a 200-Hz sound at 60 dB?

Strategy for (a)

The graph in [link] should be referenced in order to solve this example. To find the loudness of a given sound, you must know its frequency and intensity level and locate that point on the square grid, then interpolate between loudness curves to get the loudness in phons.

Solution for (a)

(1) Identify knowns:

- The square grid of the graph relating phons and decibels is a plot of intensity level versus frequency—both physical quantities.
- 100 Hz at 80 dB lies halfway between the curves marked 70 and 80 phons.

(2) Find the loudness: 75 phons.

Strategy for (b)

The graph in [link] should be referenced in order to solve this example. To find the intensity level of a sound, you must have its frequency and loudness. Once that point is located, the intensity level can be determined from the vertical axis.

Solution for (b)

(1) Identify knowns:

- Values are given to be 4000 Hz at 70 phons.

(2) Follow the 70-phon curve until it reaches 4000 Hz. At that point, it is below the 70 dB line at about 67 dB.

(3) Find the intensity level:

67 dB

Strategy for (c)

The graph in [link] should be referenced in order to solve this example.

Solution for (c)

(1) Locate the point for a 200 Hz and 60 dB sound.

(2) Find the loudness: This point lies just slightly above the 50-phon curve, and so its loudness is 51 phons.

(3) Look for the 51-phon level is at 8000 Hz: 63 dB.

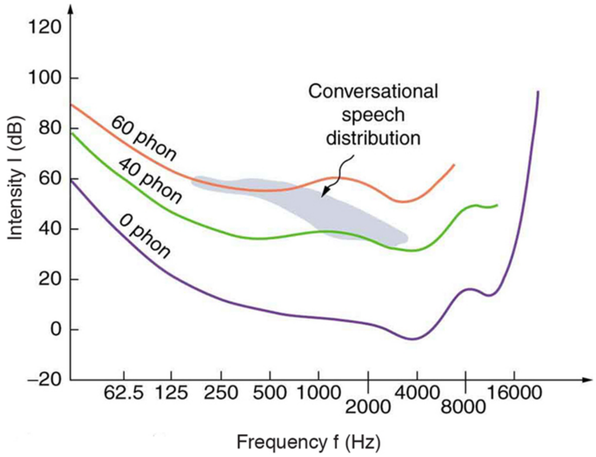
Discussion

These answers, like all information extracted from [link], have uncertainties of several phons or several decibels, partly due to difficulties in interpolation, but mostly related to uncertainties in the equal-loudness curves.

Further examination of the graph in [link] reveals some interesting facts about human hearing. First, sounds below the 0-phon curve are not perceived by most people. So, for example, a 60 Hz sound at 40 dB is inaudible. The 0-phon curve represents the threshold of normal hearing. We can hear some sounds at intensity levels below 0 dB. For example, a 3-dB, 5000-Hz sound is audible, because it lies above the 0-phon curve. The loudness curves all have dips in them between about 2000 and 5000 Hz. These dips mean the ear is most sensitive to frequencies in that range. For example, a 15-dB sound at 4000 Hz has a loudness of 20 phons, the same as a 20-dB sound at 1000 Hz. The curves rise at both extremes of the frequency range, indicating that a greater-intensity level sound is needed at those frequencies to be perceived to be as loud as at middle frequencies. For example, a sound at 10,000 Hz must have an intensity level of 30 dB to seem as loud as a 20 dB sound at 1000 Hz. Sounds above 120 phons are painful as well as damaging.

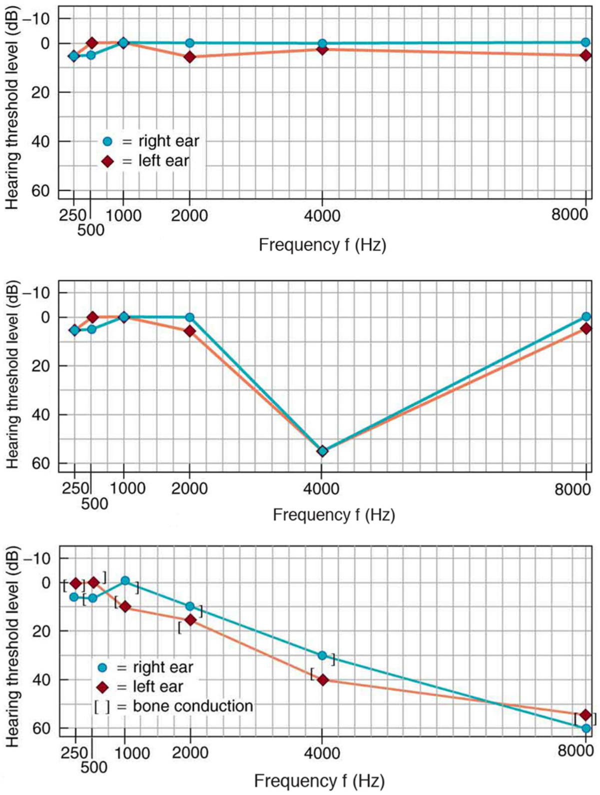
We do not often utilize our full range of hearing. This is particularly true for frequencies above 8000 Hz, which are rare in the environment and are unnecessary for understanding conversation or appreciating music. In fact, people who have lost the ability to hear such high frequencies are usually unaware of their loss until tested. The shaded region in [link] is the frequency and intensity region where most conversational sounds fall. The curved lines indicate what effect hearing losses of 40 and 60 phons will have. A 40-phon hearing loss at all frequencies still allows a person to understand conversation, although it will seem very quiet. A person with a 60-phon loss at all frequencies will hear only the lowest frequencies and will not be able to understand speech unless it is much louder than normal. Even so, speech may seem indistinct, because higher frequencies are not as well perceived. The conversational speech region also has a gender component, in that female voices are usually characterized by higher

frequencies. So the person with a 60-phon hearing impediment might have difficulty understanding the normal conversation of a woman.



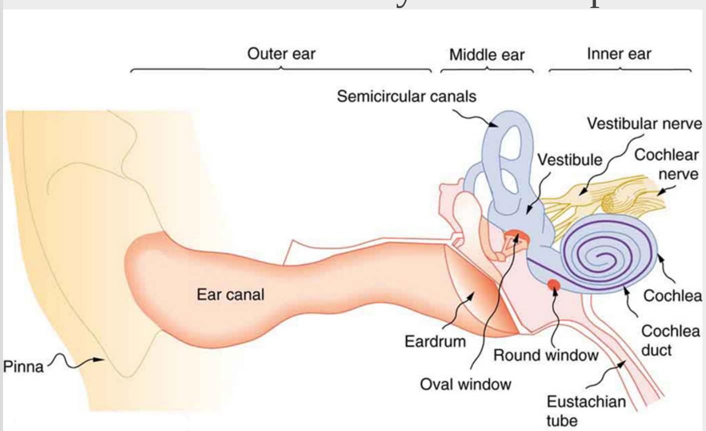
The shaded region represents frequencies and intensity levels found in normal conversational speech. The 0-phon line represents the normal hearing threshold, while those at 40 and 60 represent thresholds for people with 40- and 60-phon hearing losses, respectively.

Hearing tests are performed over a range of frequencies, usually from 250 to 8000 Hz, and can be displayed graphically in an audiogram like that in [link]. The hearing threshold is measured in dB *relative to the normal threshold*, so that normal hearing registers as 0 dB at all frequencies. Hearing loss caused by noise typically shows a dip near the 4000 Hz frequency, irrespective of the frequency that caused the loss and often affects both ears. The most common form of hearing loss comes with age and is called *presbycusis*—literally elder ear. Such loss is increasingly severe at higher frequencies, and interferes with music appreciation and speech recognition.



Audiograms showing the threshold in intensity level versus frequency for three different individuals. Intensity level is measured relative to the normal threshold. The top left graph is that of a person with normal hearing. The graph to its right has a dip at 4000 Hz and is that of a child who suffered hearing loss due to a cap gun. The third graph is typical of presbycusis, the progressive loss of higher frequency hearing with age. Tests performed by bone conduction (brackets) can distinguish nerve damage from middle ear damage.

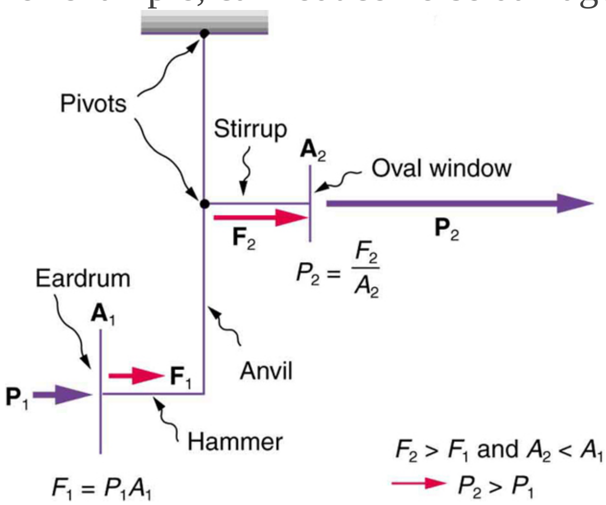
Note:
The Hearing Mechanism
The hearing mechanism involves some interesting physics. The sound wave that impinges upon our ear is a pressure wave. The ear is a transducer that converts sound waves into electrical nerve impulses in a manner much more sophisticated than, but analogous to, a microphone. [link] shows the gross anatomy of the ear with its division into three parts: the outer ear or ear canal; the middle ear, which runs from the eardrum to the cochlea; and the inner ear, which is the cochlea itself. The body part normally referred to as the ear is technically called the pinna.



The illustration shows the gross anatomy of the human ear.

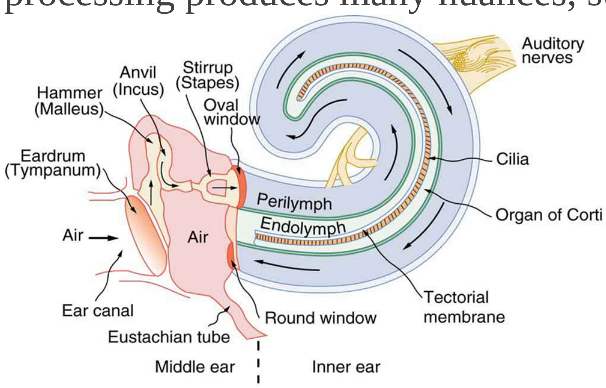
The outer ear, or ear canal, carries sound to the recessed protected eardrum. The air column in the ear canal resonates and is partially responsible for the sensitivity of the ear to sounds in the 2000 to 5000 Hz range. The middle ear converts sound into mechanical vibrations and applies these vibrations to the cochlea. The lever system of the middle ear takes the force exerted on the eardrum by sound pressure variations, amplifies it and transmits it to the

inner ear via the oval window, creating pressure waves in the cochlea approximately 40 times greater than those impinging on the eardrum. (See [link].) Two muscles in the middle ear (not shown) protect the inner ear from very intense sounds. They react to intense sound in a few milliseconds and reduce the force transmitted to the cochlea. This protective reaction can also be triggered by your own voice, so that humming while shooting a gun, for example, can reduce noise damage.



This schematic shows the middle ear's system for converting sound pressure into force, increasing that force through a lever system, and applying the increased force to a small area of the cochlea, thereby creating a pressure about 40 times that in the original sound wave. A protective muscle reaction to intense sounds greatly reduces the mechanical advantage of the lever system.

[link] shows the middle and inner ear in greater detail. Pressure waves moving through the cochlea cause the tectorial membrane to vibrate, rubbing cilia (called hair cells), which stimulate nerves that send electrical signals to the brain. The membrane resonates at different positions for different frequencies, with high frequencies stimulating nerves at the near end and low frequencies at the far end. The complete operation of the cochlea is still not understood, but several mechanisms for sending information to the brain are known to be involved. For sounds below about 1000 Hz, the nerves send signals at the same frequency as the sound. For frequencies greater than about 1000 Hz, the nerves signal frequency by position. There is a structure to the cilia, and there are connections between nerve cells that perform signal processing before information is sent to the brain. Intensity information is partly indicated by the number of nerve signals and by volleys of signals. The brain processes the cochlear nerve signals to provide additional information such as source direction (based on time and intensity comparisons of sounds from both ears). Higher-level processing produces many nuances, such as music appreciation.



The inner ear, or cochlea, is a coiled tube about 3 mm in diameter and 3 cm in length if uncoiled. When the oval window is forced inward, as shown, a pressure wave travels through the perilymph in the direction of the arrows, stimulating nerves at the base of cilia in the organ of Corti.

Hearing losses can occur because of problems in the middle or inner ear. Conductive losses in the middle ear can be partially overcome by sending sound vibrations to the cochlea through the skull. Hearing aids for this purpose usually press against the bone behind the ear, rather than simply amplifying the sound sent into the ear canal as many hearing aids do. Damage to the nerves in the cochlea is not repairable, but amplification can partially compensate. There is a risk that amplification will produce further damage. Another common failure in the cochlea is damage or loss of the cilia but with nerves remaining functional. Cochlear implants that stimulate the nerves directly are now available and widely accepted. Over 100,000 implants are in use, in about equal numbers of adults and children.

The cochlear implant was pioneered in Melbourne, Australia, by Graeme Clark in the 1970s for his deaf father. The implant consists of three external components and two internal components. The external components are a microphone for picking up sound and converting it into an electrical signal, a speech processor to select certain frequencies and a transmitter to transfer the signal to the internal components through electromagnetic induction. The internal components consist of a receiver/transmitter secured in the bone beneath the skin, which converts the signals into electric impulses and sends them through an internal cable to the cochlea and an array of about 24 electrodes wound through the cochlea. These electrodes in turn send the impulses directly into the brain. The electrodes basically emulate the cilia.

Exercise:

Check Your Understanding

Problem:

Are ultrasound and infrasound imperceptible to all hearing organisms? Explain your answer.

Solution:

No, the range of perceptible sound is based in the range of human hearing. Many other organisms perceive either infrasound or ultrasound.

Section Summary

- The range of audible frequencies is 20 to 20,000 Hz.
- Those sounds above 20,000 Hz are ultrasound, whereas those below 20 Hz are infrasound.
- The perception of frequency is pitch.
- The perception of intensity is loudness.
- Loudness has units of phons.

Conceptual Questions

Exercise:

Problem:

Why can a hearing test show that your threshold of hearing is 0 dB at 250 Hz, when [link] implies that no one can hear such a frequency at less than 20 dB?

Problems & Exercises

Exercise:

Problem:

The factor of 10^{-12} in the range of intensities to which the ear can respond, from threshold to that causing damage after brief exposure, is truly remarkable. If you could measure distances over the same range with a single instrument and the smallest distance you could measure was 1 mm, what would the largest be?

Solution:
Equation:

$$1 \times 10^{6}\ \text{km}$$

Exercise:

Problem:

The frequencies to which the ear responds vary by a factor of 10^3. Suppose the speedometer on your car measured speeds differing by the same factor of 10^3, and the greatest speed it reads is 90.0 mi/h. What would be the slowest nonzero speed it could read?

Exercise:

Problem:

What are the closest frequencies to 500 Hz that an average person can clearly distinguish as being different in frequency from 500 Hz? The sounds are not present simultaneously.

Solution:

498.5 or 501.5 Hz

Exercise:

Problem:

Can the average person tell that a 2002-Hz sound has a different frequency than a 1999-Hz sound without playing them simultaneously?

Exercise:

Problem:

If your radio is producing an average sound intensity level of 85 dB, what is the next lowest sound intensity level that is clearly less intense?

Solution:

82 dB

Exercise:

Problem:

Can you tell that your roommate turned up the sound on the TV if its average sound intensity level goes from 70 to 73 dB?

Exercise:

Problem:

Based on the graph in [link], what is the threshold of hearing in decibels for frequencies of 60, 400, 1000, 4000, and 15,000 Hz? Note that many AC electrical appliances produce 60 Hz, music is commonly 400 Hz, a reference frequency is 1000 Hz, your maximum sensitivity is near 4000 Hz, and many older TVs produce a 15,750 Hz whine.

Solution:

approximately 48, 9, 0, –7, and 20 dB, respectively

Exercise:

Problem:

What sound intensity levels must sounds of frequencies 60, 3000, and 8000 Hz have in order to have the same loudness as a 40-dB sound of frequency 1000 Hz (that is, to have a loudness of 40 phons)?

Exercise:

Problem:

What is the approximate sound intensity level in decibels of a 600-Hz tone if it has a loudness of 20 phons? If it has a loudness of 70 phons?

Solution:

(a) 23 dB

(b) 70 dB

Exercise:

Problem:

(a) What are the loudnesses in phons of sounds having frequencies of 200, 1000, 5000, and 10,000 Hz, if they are all at the same 60.0-dB sound intensity level? (b) If they are all at 110 dB? (c) If they are all at 20.0 dB?

Exercise:

Problem:

Suppose a person has a 50-dB hearing loss at all frequencies. By how many factors of 10 will low-intensity sounds need to be amplified to seem normal to this person? Note that smaller amplification is appropriate for more intense sounds to avoid further hearing damage.

Solution:

Five factors of 10

Exercise:

Problem:

If a woman needs an amplification of 5.0×10^{12} times the threshold intensity to enable her to hear at all frequencies, what is her overall hearing loss in dB? Note that smaller amplification is appropriate for more intense sounds to avoid further damage to her hearing from levels above 90 dB.

Exercise:

Problem:

(a) What is the intensity in watts per meter squared of a just barely audible 200-Hz sound? (b) What is the intensity in watts per meter squared of a barely audible 4000-Hz sound?

Solution:

(a) $2 \times 10^{-10}\ \mathrm{W/m^2}$

(b) $2 \times 10^{-13}\ \mathrm{W/m^2}$

Exercise:

Problem:

(a) Find the intensity in watts per meter squared of a 60.0-Hz sound having a loudness of 60 phons. (b) Find the intensity in watts per meter squared of a 10,000-Hz sound having a loudness of 60 phons.

Exercise:

Problem:

A person has a hearing threshold 10 dB above normal at 100 Hz and 50 dB above normal at 4000 Hz. How much more intense must a 100-Hz tone be than a 4000-Hz tone if they are both barely audible to this person?

Solution:

2.5

Exercise:

Problem:

A child has a hearing loss of 60 dB near 5000 Hz, due to noise exposure, and normal hearing elsewhere. How much more intense is a 5000-Hz tone than a 400-Hz tone if they are both barely audible to the child?

Exercise:

Problem:

What is the ratio of intensities of two sounds of identical frequency if the first is just barely discernible as louder to a person than the second?

Solution:

1.26

Glossary

loudness
: the perception of sound intensity

timbre
: number and relative intensity of multiple sound frequencies

note
: basic unit of music with specific names, combined to generate tunes

tone
: number and relative intensity of multiple sound frequencies

phon
: the numerical unit of loudness

ultrasound
: sounds above 20,000 Hz

infrasound
: sounds below 20 Hz

Ultrasound

- Define acoustic impedance and intensity reflection coefficient.
- Describe medical and other uses of ultrasound technology.
- Calculate acoustic impedance using density values and the speed of ultrasound.
- Calculate the velocity of a moving object using Doppler-shifted ultrasound.

Ultrasound is used in medicine to painlessly and noninvasively monitor patient health and diagnose a wide range of disorders. (credit: abbybatchelder, Flickr)

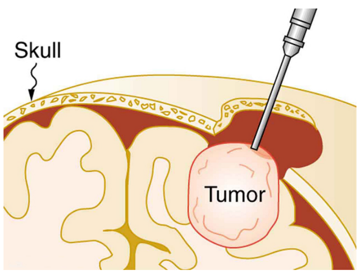
Any sound with a frequency above 20,000 Hz (or 20 kHz)—that is, above the highest audible frequency—is defined to be ultrasound. In practice, it is possible to create ultrasound frequencies up to more than a gigahertz. (Higher frequencies are difficult to create; furthermore, they propagate poorly because they are very strongly absorbed.) Ultrasound has a tremendous number of applications, which range from burglar alarms to use in cleaning delicate objects to the guidance systems of bats. We begin our discussion of ultrasound with some of its applications in medicine, in which it is used extensively both for diagnosis and for therapy.

Note:
Characteristics of Ultrasound
The characteristics of ultrasound, such as frequency and intensity, are wave properties common to all types of waves. Ultrasound also has a wavelength that limits the fineness of detail it can detect. This characteristic is true of all waves. We can never observe details significantly smaller than the wavelength of our probe; for example,

we will never see individual atoms with visible light, because the atoms are so small compared with the wavelength of light.

Ultrasound in Medical Therapy

Ultrasound, like any wave, carries energy that can be absorbed by the medium carrying it, producing effects that vary with intensity. When focused to intensities of 10^3 to $10^5\ \mathrm{W/m^2}$, ultrasound can be used to shatter gallstones or pulverize cancerous tissue in surgical procedures. (See [link].) Intensities this great can damage individual cells, variously causing their protoplasm to stream inside them, altering their permeability, or rupturing their walls through *cavitation*. Cavitation is the creation of vapor cavities in a fluid—the longitudinal vibrations in ultrasound alternatively compress and expand the medium, and at sufficient amplitudes the expansion separates molecules. Most cavitation damage is done when the cavities collapse, producing even greater shock pressures.



The tip of this small probe oscillates at 23 kHz with such a large amplitude that it pulverizes tissue on contact. The debris is then aspirated. The speed of the tip may exceed the speed of sound in tissue, thus creating shock waves and cavitation, rather than a smooth

simple harmonic
oscillator–type
wave.

Most of the energy carried by high-intensity ultrasound in tissue is converted to thermal energy. In fact, intensities of 10^3 to $10^4\ \text{W/m}^2$ are commonly used for deep-heat treatments called ultrasound diathermy. Frequencies of 0.8 to 1 MHz are typical. In both athletics and physical therapy, ultrasound diathermy is most often applied to injured or overworked muscles to relieve pain and improve flexibility. Skill is needed by the therapist to avoid “bone burns” and other tissue damage caused by overheating and cavitation, sometimes made worse by reflection and focusing of the ultrasound by joint and bone tissue.

In some instances, you may encounter a different decibel scale, called the sound *pressure* level, when ultrasound travels in water or in human and other biological tissues. We shall not use the scale here, but it is notable that numbers for sound pressure levels range 60 to 70 dB higher than you would quote for β, the sound intensity level used in this text. Should you encounter a sound pressure level of 220 decibels, then, it is not an astronomically high intensity, but equivalent to about 155 dB—high enough to destroy tissue, but not as unreasonably high as it might seem at first.

Ultrasound in Medical Diagnostics

When used for imaging, ultrasonic waves are emitted from a transducer, a crystal exhibiting the piezoelectric effect (the expansion and contraction of a substance when a voltage is applied across it, causing a vibration of the crystal). These high-frequency vibrations are transmitted into any tissue in contact with the transducer. Similarly, if a pressure is applied to the crystal (in the form of a wave reflected off tissue layers), a voltage is produced which can be recorded. The crystal therefore acts as both a transmitter and a receiver of sound. Ultrasound is also partially absorbed by tissue on its path, both on its journey away from the transducer and on its return journey. From the time between when the original signal is sent and when the reflections from various boundaries between media are received, (as well as a measure of the intensity loss of the signal), the nature and position of each boundary between tissues and organs may be deduced.

Reflections at boundaries between two different media occur because of differences in a characteristic known as the **acoustic impedance** Z of each substance. Impedance is defined as

Equation:

$$Z = \rho v,$$

where ρ is the density of the medium (in kg/m^3) and v is the speed of sound through the medium (in m/s). The units for Z are therefore $\text{kg/(m}^2 \cdot \text{s)}$.

[link] shows the density and speed of sound through various media (including various soft tissues) and the associated acoustic impedances. Note that the acoustic impedances for soft tissue do not vary much but that there is a big difference between the acoustic impedance of soft tissue and air and also between soft tissue and bone.

Medium	Density (kg/m^3)	Speed of Ultrasound (m/s)	Acoustic Impedance $\left(\text{kg/}\left(\text{m}^2 \cdot \text{s}\right)\right)$
Air	1.3	330	429
Water	1000	1500	1.5×10^6
Blood	1060	1570	1.66×10^6
Fat	925	1450	1.34×10^6
Muscle (average)	1075	1590	1.70×10^6
Bone (varies)	1400–1900	4080	5.7×10^6 to 7.8×10^6
Barium titanate (transducer material)	5600	5500	30.8×10^6

The Ultrasound Properties of Various Media, Including Soft Tissue Found in the Body

At the boundary between media of different acoustic impedances, some of the wave energy is reflected and some is transmitted. The greater the *difference* in acoustic impedance between the two media, the greater the reflection and the smaller the transmission.

The **intensity reflection coefficient** a is defined as the ratio of the intensity of the reflected wave relative to the incident (transmitted) wave. This statement can be written mathematically as
Equation:

$$a = \frac{(Z_2 - Z_1)^2}{(Z_1 + Z_2)^2},$$

where Z_1 and Z_2 are the acoustic impedances of the two media making up the boundary. A reflection coefficient of zero (corresponding to total transmission and no reflection) occurs when the acoustic impedances of the two media are the same. An impedance "match" (no reflection) provides an efficient coupling of sound energy from one medium to another. The image formed in an ultrasound is made by tracking reflections (as shown in [link]) and mapping the intensity of the reflected sound waves in a two-dimensional plane.

Example:
Calculate Acoustic Impedance and Intensity Reflection Coefficient: Ultrasound and Fat Tissue
(a) Using the values for density and the speed of ultrasound given in [link], show that the acoustic impedance of fat tissue is indeed $1.34 \times 10^6\ \text{kg/(m}^2\text{·s)}$.
(b) Calculate the intensity reflection coefficient of ultrasound when going from fat to muscle tissue.
Strategy for (a)
The acoustic impedance can be calculated using $Z = \rho v$ and the values for ρ and v found in [link].
Solution for (a)
(1) Substitute known values from [link] into $Z = \rho v$.
Equation:

$$Z = \rho v = \left(925\ \text{kg/m}^3\right)(1450\ \text{m/s})$$

(2) Calculate to find the acoustic impedance of fat tissue.
Equation:

$$1.34 \times 10^6 \text{ kg/(m}^2\text{·s)}$$

This value is the same as the value given for the acoustic impedance of fat tissue.

Strategy for (b)

The intensity reflection coefficient for any boundary between two media is given by $a = \frac{(Z_2 - Z_1)^2}{(Z_1 + Z_2)^2}$, and the acoustic impedance of muscle is given in [link].

Solution for (b)

Substitute known values into $a = \frac{(Z_2 - Z_1)^2}{(Z_1 + Z_2)^2}$ to find the intensity reflection coefficient:

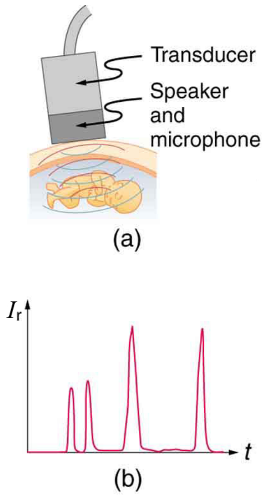
Equation:

$$a = \frac{(Z_2 - Z_1)^2}{(Z_1 + Z_2)^2} = \frac{\left(1.34 \times 10^6 \text{ kg/(m}^2\text{· s)} - 1.70 \times 10^6 \text{ kg/(m}^2\text{· s)}\right)^2}{\left(1.70 \times 10^6 \text{ kg/(m}^2\text{· s)} + 1.34 \times 10^6 \text{ kg/(m}^2\text{· s)}\right)^2} = 0.014$$

Discussion

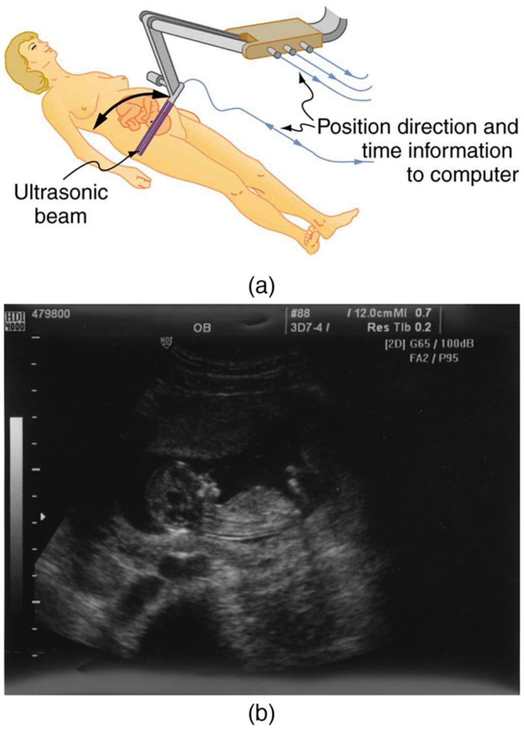
This result means that only 1.4% of the incident intensity is reflected, with the remaining being transmitted.

The applications of ultrasound in medical diagnostics have produced untold benefits with no known risks. Diagnostic intensities are too low (about 10^{-2} W/m^2) to cause thermal damage. More significantly, ultrasound has been in use for several decades and detailed follow-up studies do not show evidence of ill effects, quite unlike the case for x-rays.



(a) An ultrasound speaker doubles as a microphone. Brief bleeps are broadcast, and echoes are recorded from various depths. (b) Graph of echo intensity versus time. The time for echoes to return is directly proportional to the distance of the reflector, yielding this information noninvasively.

The most common ultrasound applications produce an image like that shown in [link]. The speaker-microphone broadcasts a directional beam, sweeping the beam across the area of interest. This is accomplished by having multiple ultrasound sources in the probe's head, which are phased to interfere constructively in a given, adjustable direction. Echoes are measured as a function of position as well as depth. A computer constructs an image that reveals the shape and density of internal structures.



(a) An ultrasonic image is produced by sweeping the ultrasonic beam across the area of interest, in this case the woman's abdomen. Data are recorded and analyzed in a computer, providing a two-dimensional image. (b) Ultrasound image of 12-week-old fetus. (credit: Margaret W. Carruthers, Flickr)

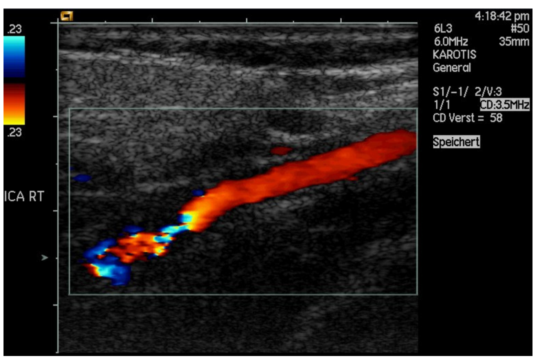
How much detail can ultrasound reveal? The image in [link] is typical of low-cost systems, but that in [link] shows the remarkable detail possible with more advanced systems, including 3D imaging. Ultrasound today is commonly used in prenatal care. Such imaging can be used to see if the fetus is developing at a normal rate, and help in the determination of serious problems early in the pregnancy. Ultrasound is also in wide use to image the chambers of the heart and the flow of blood within the beating heart, using the Doppler effect (echocardiology).

Whenever a wave is used as a probe, it is very difficult to detect details smaller than its wavelength λ. Indeed, current technology cannot do quite this well. Abdominal scans may use a 7-MHz frequency, and the speed of sound in tissue is about 1540 m/s —so the wavelength limit to detail would be $\lambda = \frac{v_{\mathrm{w}}}{f} = \frac{1540\ \mathrm{m/s}}{7 \times 10^6\ \mathrm{Hz}} = 0.22\ \mathrm{mm}$. In practice, 1-mm detail is attainable, which is sufficient for many purposes. Higher-frequency ultrasound would allow greater detail, but it does not penetrate as well as lower frequencies do. The accepted rule of thumb is that you can effectively scan to a depth of about 500λ into tissue. For 7 MHz, this penetration limit is $500 \times 0.22\ \mathrm{mm}$, which is 0.11 m. Higher frequencies may be employed in smaller organs, such as the eye, but are not practical for looking deep into the body.

A 3D ultrasound image of a fetus. As well as for the detection of any abnormalities, such scans have also been shown to be useful for strengthening the emotional bonding between parents and their unborn child. (credit: Jennie Cu, Wikimedia Commons)

In addition to shape information, ultrasonic scans can produce density information superior to that found in X-rays, because the intensity of a reflected sound is related to changes in density. Sound is most strongly reflected at places where density changes are greatest.

Another major use of ultrasound in medical diagnostics is to detect motion and determine velocity through the Doppler shift of an echo, known as **Doppler-shifted ultrasound**. This technique is used to monitor fetal heartbeat, measure blood velocity, and detect occlusions in blood vessels, for example. (See [link].) The magnitude of the Doppler shift in an echo is directly proportional to the velocity of whatever reflects the sound. Because an echo is involved, there is actually a double shift. The first occurs because the reflector (say a fetal heart) is a moving observer and receives a Doppler-shifted frequency. The reflector then acts as a moving source, producing a second Doppler shift.



This Doppler-shifted ultrasonic image of a partially occluded artery uses color to indicate velocity. The highest velocities are in red, while the lowest are blue. The blood must move faster through the constriction to carry the same flow. (credit: Arning C, Grzyska U, Wikimedia Commons)

A clever technique is used to measure the Doppler shift in an echo. The frequency of the echoed sound is superimposed on the broadcast frequency, producing beats. The beat frequency is $F_{\text{B}} = \mid f_1 - f_2 \mid$, and so it is directly proportional to the Doppler shift $(f_1 - f_2)$ and hence, the reflector's velocity. The advantage in this technique is that the Doppler shift is small (because the reflector's velocity is small), so that great accuracy would be needed to measure the shift directly. But measuring the beat frequency is easy, and it is not affected if the broadcast frequency varies somewhat. Furthermore, the beat frequency is in the audible range and can be amplified for audio feedback to the medical observer.

Note:

Uses for Doppler-Shifted Radar

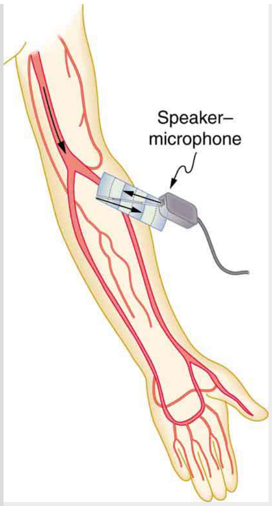
Doppler-shifted radar echoes are used to measure wind velocities in storms as well as aircraft and automobile speeds. The principle is the same as for Doppler-shifted ultrasound. There is evidence that bats and dolphins may also sense the velocity of an object (such as prey) reflecting their ultrasound signals by observing its Doppler shift.

Example:

Calculate Velocity of Blood: Doppler-Shifted Ultrasound

Ultrasound that has a frequency of 2.50 MHz is sent toward blood in an artery that is moving toward the source at 20.0 cm/s, as illustrated in [link]. Use the speed of sound in human tissue as 1540 m/s. (Assume that the frequency of 2.50 MHz is accurate to seven significant figures.)

a. What frequency does the blood receive?
b. What frequency returns to the source?
c. What beat frequency is produced if the source and returning frequencies are mixed?



Ultrasound is partly reflected by blood cells and plasma back toward the speaker-microphone. Because the cells are moving, two Doppler shifts are produced—one for blood as a moving observer, and the other for the reflected sound coming from a moving source. The magnitude of the shift is directly proportional to blood velocity.

Strategy

The first two questions can be answered using $f_{\text{obs}} = f_{\text{s}}\left(\frac{v_{\text{w}}}{v_{\text{w}} \pm v_{\text{s}}}\right)$ and $f_{\text{obs}} = f_{\text{s}}\left(\frac{v_{\text{w}} \pm v_{\text{obs}}}{v_{\text{w}}}\right)$ for the Doppler shift. The last question asks for beat frequency, which is the difference between the original and returning frequencies.

Solution for (a)

(1) Identify knowns:

- The blood is a moving observer, and so the frequency it receives is given by
  Equation:

$$f_{\text{obs}} = f_{\text{s}}\left(\frac{v_{\text{w}} \pm v_{\text{obs}}}{v_{\text{w}}}\right).$$

- v_{b} is the blood velocity (v_{obs} here) and the plus sign is chosen because the motion is toward the source.

(2) Enter the given values into the equation.

Equation:

$$f_{\text{obs}} = (2{,}500{,}000\ \text{Hz})\left(\frac{1540\ \text{m/s} + 0.2\ \text{m/s}}{1540\ \text{m/s}}\right)$$

(3) Calculate to find the frequency: 2,500,325 Hz.

Solution for (b)

(1) Identify knowns:

- The blood acts as a moving source.
- The microphone acts as a stationary observer.
- The frequency leaving the blood is 2,500,325 Hz, but it is shifted upward as given by
  Equation:

$$f_{\text{obs}} = f_{\text{s}}\left(\frac{v_{\text{w}}}{v_{\text{w}} - v_{\text{b}}}\right).$$

  f_{obs} is the frequency received by the speaker-microphone.
- The source velocity is v_{b}.
- The minus sign is used because the motion is toward the observer.

The minus sign is used because the motion is toward the observer.
(2) Enter the given values into the equation:
Equation:

$$f_{\text{obs}} = (2{,}500{,}325\text{ Hz})\left(\frac{1540\text{ m/s}}{1540\text{ m/s} - 0.200\text{ m/s}}\right)$$

(3) Calculate to find the frequency returning to the source: 2,500,649 Hz.
Solution for (c)
(1) Identify knowns:

- The beat frequency is simply the absolute value of the difference between f_{s} and f_{obs}, as stated in:
  Equation:

  $$f_{\text{B}} = \mid f_{\text{obs}} - f_{\text{s}} \mid.$$

(2) Substitute known values:
Equation:

$$\mid 2{,}500{,}649\text{ Hz} - 2{,}500{,}000\text{ Hz} \mid$$

(3) Calculate to find the beat frequency: 649 Hz.
Discussion
The Doppler shifts are quite small compared with the original frequency of 2.50 MHz. It is far easier to measure the beat frequency than it is to measure the echo frequency with an accuracy great enough to see shifts of a few hundred hertz out of a couple of megahertz. Furthermore, variations in the source frequency do not greatly affect the beat frequency, because both f_{s} and f_{obs}would increase or decrease. Those changes subtract out in $f_{\text{B}} = \mid f_{\text{obs}} - f_{\text{s}} \mid$.

Note:
Industrial and Other Applications of Ultrasound
Industrial, retail, and research applications of ultrasound are common. A few are discussed here. Ultrasonic cleaners have many uses. Jewelry, machined parts, and other objects that have odd shapes and crevices are immersed in a cleaning fluid that is agitated with ultrasound typically about 40 kHz in frequency. The intensity is great enough to cause cavitation, which is responsible for most of the cleansing action. Because cavitation-produced shock pressures are large and well transmitted in a fluid,

they reach into small crevices where even a low-surface-tension cleaning fluid might not penetrate.

Sonar is a familiar application of ultrasound. Sonar typically employs ultrasonic frequencies in the range from 30.0 to 100 kHz. Bats, dolphins, submarines, and even some birds use ultrasonic sonar. Echoes are analyzed to give distance and size information both for guidance and finding prey. In most sonar applications, the sound reflects quite well because the objects of interest have significantly different density than the medium in which they travel. When the Doppler shift is observed, velocity information can also be obtained. Submarine sonar can be used to obtain such information, and there is evidence that some bats also sense velocity from their echoes.

Similarly, there are a range of relatively inexpensive devices that measure distance by timing ultrasonic echoes. Many cameras, for example, use such information to focus automatically. Some doors open when their ultrasonic ranging devices detect a nearby object, and certain home security lights turn on when their ultrasonic rangers observe motion. Ultrasonic “measuring tapes” also exist to measure such things as room dimensions. Sinks in public restrooms are sometimes automated with ultrasound devices to turn faucets on and off when people wash their hands. These devices reduce the spread of germs and can conserve water.

Ultrasound is used for nondestructive testing in industry and by the military. Because ultrasound reflects well from any large change in density, it can reveal cracks and voids in solids, such as aircraft wings, that are too small to be seen with x-rays. For similar reasons, ultrasound is also good for measuring the thickness of coatings, particularly where there are several layers involved.

Basic research in solid state physics employs ultrasound. Its attenuation is related to a number of physical characteristics, making it a useful probe. Among these characteristics are structural changes such as those found in liquid crystals, the transition of a material to a superconducting phase, as well as density and other properties.

These examples of the uses of ultrasound are meant to whet the appetites of the curious, as well as to illustrate the underlying physics of ultrasound. There are many more applications, as you can easily discover for yourself.

Exercise:
Check Your Understanding

Problem:

Why is it possible to use ultrasound both to observe a fetus in the womb and also to destroy cancerous tumors in the body?

Solution:

Ultrasound can be used medically at different intensities. Lower intensities do not cause damage and are used for medical imaging. Higher intensities can pulverize and destroy targeted substances in the body, such as tumors.

Section Summary

- The acoustic impedance is defined as:
  Equation:
  $$Z = \rho v,$$
  ρ is the density of a medium through which the sound travels and v is the speed of sound through that medium.
- The intensity reflection coefficient a, a measure of the ratio of the intensity of the wave reflected off a boundary between two media relative to the intensity of the incident wave, is given by
  Equation:
  $$a = \frac{(Z_2 - Z_1)^2}{(Z_1 + Z_2)^2}.$$
- The intensity reflection coefficient is a unitless quantity.

Conceptual Questions

Exercise:

Problem:

If audible sound follows a rule of thumb similar to that for ultrasound, in terms of its absorption, would you expect the high or low frequencies from your neighbor’s stereo to penetrate into your house? How does this expectation compare with your experience?

Exercise:

Problem:

Elephants and whales are known to use infrasound to communicate over very large distances. What are the advantages of infrasound for long distance communication?

Exercise:

Problem:

It is more difficult to obtain a high-resolution ultrasound image in the abdominal region of someone who is overweight than for someone who has a slight build. Explain why this statement is accurate.

Exercise:

Problem:

Suppose you read that 210-dB ultrasound is being used to pulverize cancerous tumors. You calculate the intensity in watts per centimeter squared and find it is unreasonably high (10^5 W/cm^2). What is a possible explanation?

Problems & Exercises

Unless otherwise indicated, for problems in this section, assume that the speed of sound through human tissues is 1540 m/s.

Exercise:

Problem:

What is the sound intensity level in decibels of ultrasound of intensity 10^5 W/m^2, used to pulverize tissue during surgery?

Solution:

170 dB

Exercise:

Problem:

Is 155-dB ultrasound in the range of intensities used for deep heating? Calculate the intensity of this ultrasound and compare this intensity with values quoted in the text.

Exercise:

Problem:

Find the sound intensity level in decibels of $2.00 \times 10^{-2} \text{ W/m}^2$ ultrasound used in medical diagnostics.

Solution:

103 dB

Exercise:

Problem:

The time delay between transmission and the arrival of the reflected wave of a signal using ultrasound traveling through a piece of fat tissue was 0.13 ms. At what depth did this reflection occur?

Exercise:

Problem:

In the clinical use of ultrasound, transducers are always coupled to the skin by a thin layer of gel or oil, replacing the air that would otherwise exist between the transducer and the skin. (a) Using the values of acoustic impedance given in [link] calculate the intensity reflection coefficient between transducer material and air. (b) Calculate the intensity reflection coefficient between transducer material and gel (assuming for this problem that its acoustic impedance is identical to that of water). (c) Based on the results of your calculations, explain why the gel is used.

Solution:

(a) 1.00

(b) 0.823

(c) Gel is used to facilitate the transmission of the ultrasound between the transducer and the patient’s body.

Exercise:

Problem:

(a) Calculate the minimum frequency of ultrasound that will allow you to see details as small as 0.250 mm in human tissue. (b) What is the effective depth to which this sound is effective as a diagnostic probe?

Exercise:

Problem:

(a) Find the size of the smallest detail observable in human tissue with 20.0-MHz ultrasound. (b) Is its effective penetration depth great enough to examine the entire eye (about 3.00 cm is needed)? (c) What is the wavelength of such ultrasound in $0^{\circ}\mathrm{C}$ air?

Solution:

(a) $77.0\ \mu\mathrm{m}$

(b) Effective penetration depth = 3.85 cm, which is enough to examine the eye.

(c) $16.6\ \mu\mathrm{m}$

Exercise:

Problem:

(a) Echo times are measured by diagnostic ultrasound scanners to determine distances to reflecting surfaces in a patient. What is the difference in echo times for tissues that are 3.50 and 3.60 cm beneath the surface? (This difference is the minimum resolving time for the scanner to see details as small as 0.100 cm, or 1.00 mm. Discrimination of smaller time differences is needed to see smaller details.) (b) Discuss whether the period T of this ultrasound must be smaller than the minimum time resolution. If so, what is the minimum frequency of the ultrasound and is that out of the normal range for diagnostic ultrasound?

Exercise:

Problem:

(a) How far apart are two layers of tissue that produce echoes having round-trip times (used to measure distances) that differ by $0.750\ \mu\mathrm{s}$? (b) What minimum frequency must the ultrasound have to see detail this small?

Solution:

(a) $5.78 \times 10^{-4}\ \mathrm{m}$

(b) $2.67 \times 10^{6}\ \mathrm{Hz}$

Exercise:

Problem:

(a) A bat uses ultrasound to find its way among trees. If this bat can detect echoes 1.00 ms apart, what minimum distance between objects can it detect? (b) Could this distance explain the difficulty that bats have finding an open door when they accidentally get into a house?

Exercise:

Problem:

A dolphin is able to tell in the dark that the ultrasound echoes received from two sharks come from two different objects only if the sharks are separated by 3.50 m, one being that much farther away than the other. (a) If the ultrasound has a frequency of 100 kHz, show this ability is not limited by its wavelength. (b) If this ability is due to the dolphin's ability to detect the arrival times of echoes, what is the minimum time difference the dolphin can perceive?

Solution:

(a) $v_{\text{w}} = 1540\text{ m/s} = f\lambda \Rightarrow \lambda = \frac{1540\text{ m/s}}{100 \times 10^3\text{ Hz}} = 0.0154\text{ m} < 3.50\text{ m}$. Because the wavelength is much shorter than the distance in question, the wavelength is not the limiting factor.

(b) 4.55 ms

Exercise:

Problem:

A diagnostic ultrasound echo is reflected from moving blood and returns with a frequency 500 Hz higher than its original 2.00 MHz. What is the velocity of the blood? (Assume that the frequency of 2.00 MHz is accurate to seven significant figures and 500 Hz is accurate to three significant figures.)

Exercise:

Problem:

Ultrasound reflected from an oncoming bloodstream that is moving at 30.0 cm/s is mixed with the original frequency of 2.50 MHz to produce beats. What is the beat frequency? (Assume that the frequency of 2.50 MHz is accurate to seven significant figures.)

Solution:

974 Hz

(Note: extra digits were retained in order to show the difference.)

Glossary

acoustic impedance
: property of medium that makes the propagation of sound waves more difficult

intensity reflection coefficient
: a measure of the ratio of the intensity of the wave reflected off a boundary between two media relative to the intensity of the incident wave

Doppler-shifted ultrasound
: a medical technique to detect motion and determine velocity through the Doppler shift of an echo

Introduction to Magnetism
class="introduction"

The magnificent spectacle of the Aurora Borealis, or northern lights, glows in the northern sky above Bear Lake near Eielson Air Force Base, Alaska. Shaped by the Earth’s magnetic field, this light is produced by radiation spewed from solar storms. (credit: Senior Airman Joshua Strang, via Flickr)

One evening, an Alaskan sticks a note to his refrigerator with a small magnet. Through the kitchen window, the Aurora Borealis glows in the night sky. This grand spectacle is shaped by the same force that holds the note to the refrigerator.

People have been aware of magnets and magnetism for thousands of years. The earliest records date to well before the time of Christ, particularly in a region of Asia Minor called Magnesia (the name of this region is the source of words like *magnetic*). Magnetic rocks found in Magnesia, which is now part of western Turkey, stimulated interest during ancient times. A practical application for magnets was found later, when they were employed as navigational compasses. The use of magnets in compasses resulted not only in improved long-distance sailing, but also in the names of “north” and “south” being given to the two types of magnetic poles.

Today magnetism plays many important roles in our lives. Physicists’ understanding of magnetism has enabled the development of technologies that affect our everyday lives. The iPod in your purse or backpack, for example, wouldn’t have been possible without the applications of magnetism and electricity on a small scale.

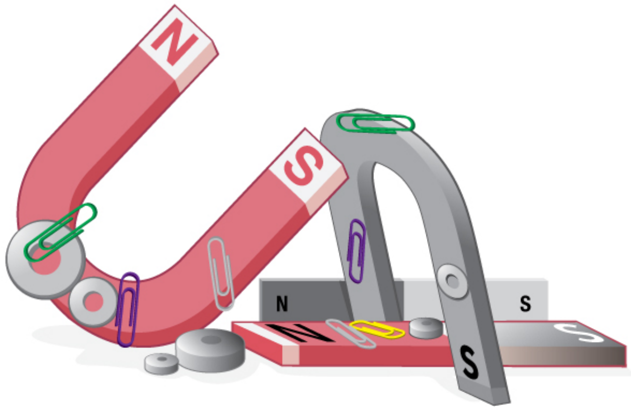
The discovery that weak changes in a magnetic field in a thin film of iron and chromium could bring about much larger changes in electrical resistance was one of the first large successes of nanotechnology. The 2007 Nobel Prize in Physics went to Albert Fert from France and Peter Grunberg from Germany for this discovery of *giant magnetoresistance* and its applications to computer memory.

All electric motors, with uses as diverse as powering refrigerators, starting cars, and moving elevators, contain magnets. Generators, whether producing hydroelectric power or running bicycle lights, use magnetic fields. Recycling facilities employ magnets to separate iron from other refuse. Hundreds of millions of dollars are spent annually on magnetic containment of fusion as a future energy source. Magnetic resonance imaging (MRI) has become an important diagnostic tool in the field of medicine, and the use of magnetism to explore brain activity is a subject of contemporary research and development. The list of applications also includes computer hard drives, tape recording, detection of inhaled asbestos, and levitation of high-speed trains. Magnetism is used to explain atomic energy levels, cosmic rays, and charged particles trapped in the Van Allen belts. Once again, we will find all these disparate phenomena are linked by a small number of underlying physical principles.

Engineering of technology like iPods would not be possible without a deep understanding magnetism. (credit: Jesse! S?, Flickr)

Magnets

- Describe the difference between the north and south poles of a magnet.
- Describe how magnetic poles interact with each other.

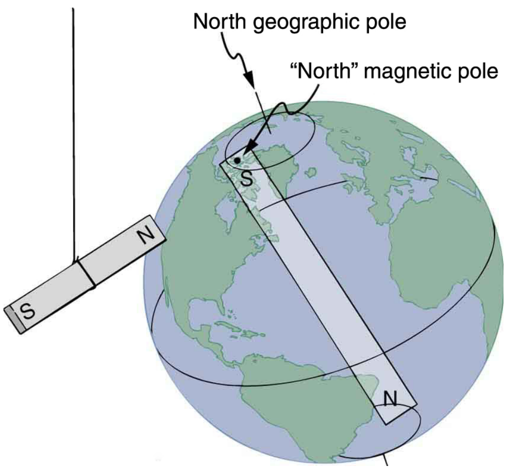


Magnets come in various shapes, sizes, and strengths. All have both a north pole and a south pole. There is never an isolated pole (a monopole).

All magnets attract iron, such as that in a refrigerator door. However, magnets may attract or repel other magnets. Experimentation shows that all magnets have two poles. If freely suspended, one pole will point toward the north. The two poles are thus named the **north magnetic pole** and the **south magnetic pole** (or more properly, north-seeking and south-seeking poles, for the attractions in those directions).

Note:
Universal Characteristics of Magnets and Magnetic Poles
It is a universal characteristic of all magnets that *like poles repel and unlike poles attract*. (Note the similarity with electrostatics: unlike charges attract and like charges repel.)

Further experimentation shows that it is *impossible to separate north and south poles* in the manner that + and − charges can be separated.



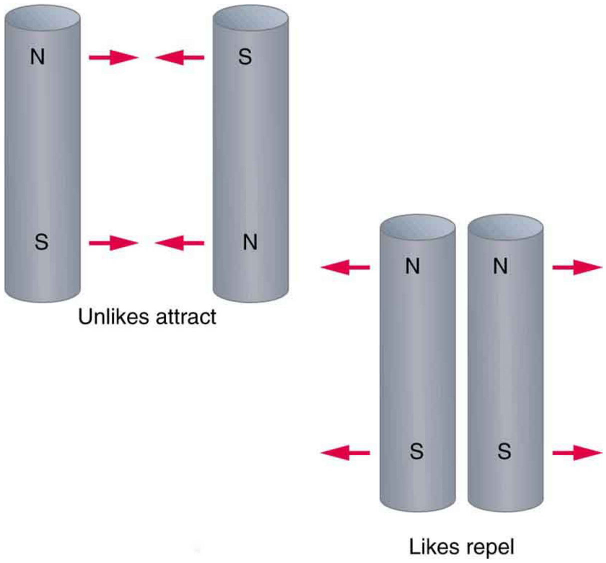
One end of a bar magnet is suspended from a thread that points toward north. The magnet's two poles are labeled N and S for north-seeking and south-seeking poles, respectively.

Note:

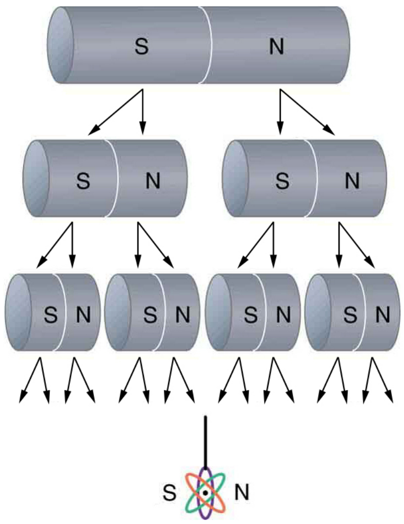
Misconception Alert: Earth's Geographic North Pole Hides an S

The Earth acts like a very large bar magnet with its south-seeking pole near the geographic North Pole. That is why the north pole of your compass is attracted toward the geographic north pole of the Earth—because the magnetic pole that is near the geographic North Pole is actually a south magnetic pole! Confusion arises because the geographic term "North Pole" has come to be used (incorrectly) for the magnetic pole that is near the

North Pole. Thus, “North magnetic pole” is actually a misnomer—it should be called the South magnetic pole.



Unlike poles attract, whereas like poles repel.



North and south poles always occur in pairs. Attempts

to separate them result in more pairs of poles. If we continue to split the magnet, we will eventually get down to an iron atom with a north pole and a south pole—these, too, cannot be separated.

The fact that magnetic poles always occur in pairs of north and south is true from the very large scale—for example, sunspots always occur in pairs that are north and south magnetic poles—all the way down to the very small scale. Magnetic atoms have both a north pole and a south pole, as do many types of subatomic particles, such as electrons, protons, and neutrons.

Note:
Making Connections: Take-Home Experiment—Refrigerator Magnets
We know that like magnetic poles repel and unlike poles attract. See if you can show this for two refrigerator magnets. Will the magnets stick if you turn them over? Why do they stick to the door anyway? What can you say about the magnetic properties of the door next to the magnet? Do refrigerator magnets stick to metal or plastic spoons? Do they stick to all types of metal?

Section Summary

- Magnetism is a subject that includes the properties of magnets, the effect of the magnetic force on moving charges and currents, and the

creation of magnetic fields by currents.
- There are two types of magnetic poles, called the north magnetic pole and south magnetic pole.
- North magnetic poles are those that are attracted toward the Earth's geographic north pole.
- Like poles repel and unlike poles attract.
- Magnetic poles always occur in pairs of north and south—it is not possible to isolate north and south poles.

Conceptual Questions

Exercise:

Problem:

Volcanic and other such activity at the mid-Atlantic ridge extrudes material to fill the gap between separating tectonic plates associated with continental drift. The magnetization of rocks is found to reverse in a coordinated manner with distance from the ridge. What does this imply about the Earth's magnetic field and how could the knowledge of the spreading rate be used to give its historical record?

Glossary

north magnetic pole
: the end or the side of a magnet that is attracted toward Earth's geographic north pole

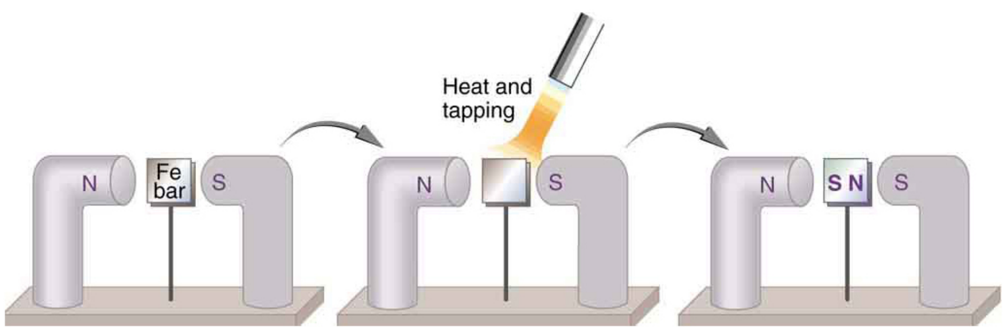
south magnetic pole
: the end or the side of a magnet that is attracted toward Earth's geographic south pole

Ferromagnets and Electromagnets

- Define ferromagnet.
- Describe the role of magnetic domains in magnetization.
- Explain the significance of the Curie temperature.
- Describe the relationship between electricity and magnetism.

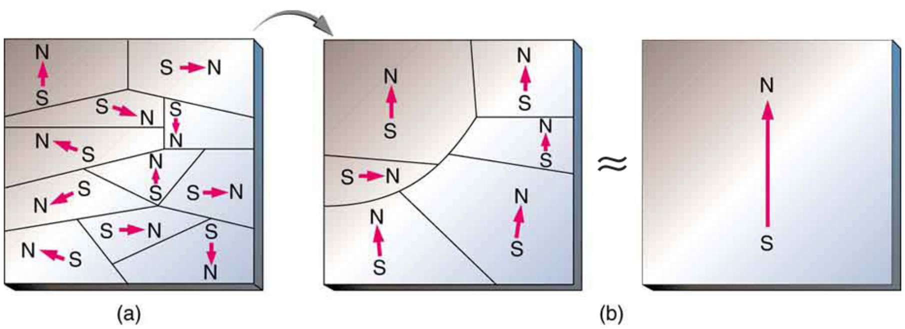
Ferromagnets

Only certain materials, such as iron, cobalt, nickel, and gadolinium, exhibit strong magnetic effects. Such materials are called **ferromagnetic**, after the Latin word for iron, *ferrum*. A group of materials made from the alloys of the rare earth elements are also used as strong and permanent magnets; a popular one is neodymium. Other materials exhibit weak magnetic effects, which are detectable only with sensitive instruments. Not only do ferromagnetic materials respond strongly to magnets (the way iron is attracted to magnets), they can also be **magnetized** themselves—that is, they can be induced to be magnetic or made into permanent magnets.



An unmagnetized piece of iron is placed between two magnets, heated, and then cooled, or simply tapped when cold. The iron becomes a permanent magnet with the poles aligned as shown: its south pole is adjacent to the north pole of the original magnet, and its north pole is adjacent to the south pole of the original magnet. Note that there are attractive forces between the magnets.

When a magnet is brought near a previously unmagnetized ferromagnetic material, it causes local magnetization of the material with unlike poles closest, as in [link]. (This results in the attraction of the previously unmagnetized material to the magnet.) What happens on a microscopic scale is illustrated in [link]. The regions within the material called **domains** act like small bar magnets. Within domains, the poles of individual atoms are aligned. Each atom acts like a tiny bar magnet. Domains are small and randomly oriented in an unmagnetized ferromagnetic object. In response to an external magnetic field, the domains may grow to millimeter size, aligning themselves as shown in [link](b). This induced magnetization can be made permanent if the material is heated and then cooled, or simply tapped in the presence of other magnets.



(a) An unmagnetized piece of iron (or other ferromagnetic material) has randomly oriented domains. (b) When magnetized by an external field, the domains show greater alignment, and some grow at the expense of others. Individual atoms are aligned within domains; each atom acts like a tiny bar magnet.

Conversely, a permanent magnet can be demagnetized by hard blows or by heating it in the absence of another magnet. Increased thermal motion at higher temperature can disrupt and randomize the orientation and the size of

the domains. There is a well-defined temperature for ferromagnetic materials, which is called the **Curie temperature**, above which they cannot be magnetized. The Curie temperature for iron is 1043 K $(770^{\circ}C)$, which is well above room temperature. There are several elements and alloys that have Curie temperatures much lower than room temperature and are ferromagnetic only below those temperatures.

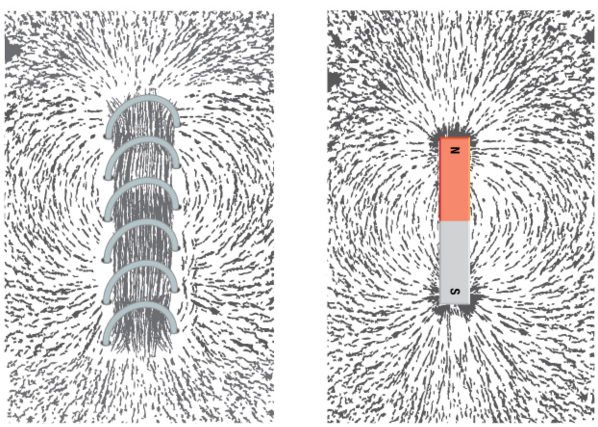
Electromagnets

Early in the 19th century, it was discovered that electrical currents cause magnetic effects. The first significant observation was by the Danish scientist Hans Christian Oersted (1777–1851), who found that a compass needle was deflected by a current-carrying wire. This was the first significant evidence that the movement of charges had any connection with magnets. **Electromagnetism** is the use of electric current to make magnets. These temporarily induced magnets are called **electromagnets**. Electromagnets are employed for everything from a wrecking yard crane that lifts scrapped cars to controlling the beam of a 90-km-circumference particle accelerator to the magnets in medical imaging machines (See [link]).

Instrument for magnetic resonance imaging (MRI). The device uses a superconducting

cylindrical coil for the main magnetic field. The patient goes into this “tunnel” on the gurney. (credit: Bill McChesney, Flickr)

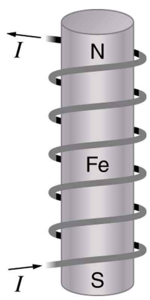
[link] shows that the response of iron filings to a current-carrying coil and to a permanent bar magnet. The patterns are similar. In fact, electromagnets and ferromagnets have the same basic characteristics—for example, they have north and south poles that cannot be separated and for which like poles repel and unlike poles attract.



(a) (b)

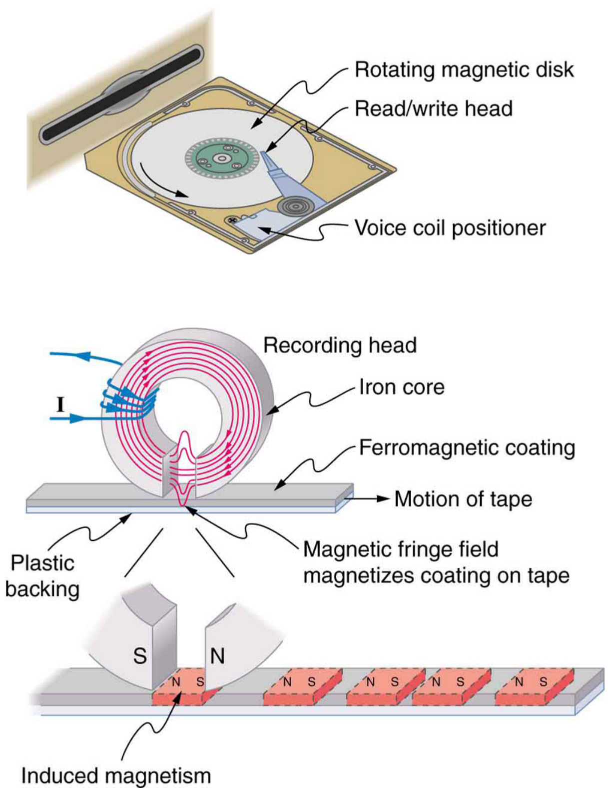
Iron filings near (a) a current-carrying coil and (b) a magnet act like tiny compass needles, showing the shape of their fields. Their response to a current-carrying coil and a permanent magnet is seen to be very similar, especially near the ends of the coil and the magnet.

Combining a ferromagnet with an electromagnet can produce particularly strong magnetic effects. (See [link].) Whenever strong magnetic effects are needed, such as lifting scrap metal, or in particle accelerators, electromagnets are enhanced by ferromagnetic materials. Limits to how strong the magnets can be made are imposed by coil resistance (it will overheat and melt at sufficiently high current), and so superconducting magnets may be employed. These are still limited, because superconducting properties are destroyed by too great a magnetic field.



An electromagnet with a ferromagnetic core can produce very strong magnetic effects. Alignment of domains in the core produces a magnet, the poles of which are aligned with the electromagnet.

[link] shows a few uses of combinations of electromagnets and ferromagnets. Ferromagnetic materials can act as memory devices, because the orientation of the magnetic fields of small domains can be reversed or erased. Magnetic information storage on videotapes and computer hard drives are among the most common applications. This property is vital in our digital world.



An electromagnet induces regions of permanent magnetism on a floppy disk coated with a ferromagnetic material. The information stored here is digital (a region is either magnetic or not); in other applications, it can be analog

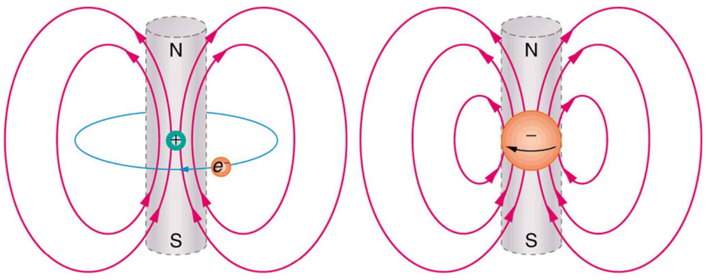
(with a varying strength), such as on audiotapes.

Current: The Source of All Magnetism

An electromagnet creates magnetism with an electric current. In later sections we explore this more quantitatively, finding the strength and direction of magnetic fields created by various currents. But what about ferromagnets? [link] shows models of how electric currents create magnetism at the submicroscopic level. (Note that we cannot directly observe the paths of individual electrons about atoms, and so a model or visual image, consistent with all direct observations, is made. We can directly observe the electron's orbital angular momentum, its spin momentum, and subsequent magnetic moments, all of which are explained with electric-current-creating subatomic magnetism.) Currents, including those associated with other submicroscopic particles like protons, allow us to explain ferromagnetism and all other magnetic effects. Ferromagnetism, for example, results from an internal cooperative alignment of electron spins, possible in some materials but not in others.

Crucial to the statement that electric current is the source of all magnetism is the fact that it is impossible to separate north and south magnetic poles. (This is far different from the case of positive and negative charges, which are easily separated.) A current loop always produces a magnetic dipole—that is, a magnetic field that acts like a north pole and south pole pair. Since isolated north and south magnetic poles, called **magnetic monopoles**, are not observed, currents are used to explain all magnetic effects. If magnetic monopoles did exist, then we would have to modify this underlying connection that all magnetism is due to electrical current. There is no known reason that magnetic monopoles should not exist—they are simply never observed—and so searches at the subnuclear level continue. If they do *not* exist, we would like to find out why not. If they *do* exist, we would like to see evidence of them.

Note:
Electric Currents and Magnetism
Electric current is the source of all magnetism.



(a) In the planetary model of the atom, an electron orbits a nucleus, forming a closed-current loop and producing a magnetic field with a north pole and a south pole. (b) Electrons have spin and can be crudely pictured as rotating charge, forming a current that produces a magnetic field with a north pole and a south pole. Neither the planetary model nor the image of a spinning electron is completely consistent with modern physics. However, they do provide a useful way of understanding phenomena.

Note:
PhET Explorations: Magnets and Electromagnets
Explore the interactions between a compass and bar magnet. Discover how you can use a battery and wire to make a magnet! Can you make it a stronger magnet? Can you make the magnetic field reverse?

https://archive.cnx.org/specials/92176000-ae74-11e5-baad-cfab91c15075/magnets-and-electromagnets/#sim-bar-magnet

Section Summary

- Magnetic poles always occur in pairs of north and south—it is not possible to isolate north and south poles.
- All magnetism is created by electric current.
- Ferromagnetic materials, such as iron, are those that exhibit strong magnetic effects.
- The atoms in ferromagnetic materials act like small magnets (due to currents within the atoms) and can be aligned, usually in millimeter-sized regions called domains.
- Domains can grow and align on a larger scale, producing permanent magnets. Such a material is magnetized, or induced to be magnetic.
- Above a material's Curie temperature, thermal agitation destroys the alignment of atoms, and ferromagnetism disappears.
- Electromagnets employ electric currents to make magnetic fields, often aided by induced fields in ferromagnetic materials.

Glossary

ferromagnetic
: materials, such as iron, cobalt, nickel, and gadolinium, that exhibit strong magnetic effects

magnetized
: to be turned into a magnet; to be induced to be magnetic

domains
: regions within a material that behave like small bar magnets

Curie temperature
: the temperature above which a ferromagnetic material cannot be magnetized

electromagnetism

the use of electrical currents to induce magnetism

electromagnet

an object that is temporarily magnetic when an electrical current is passed through it

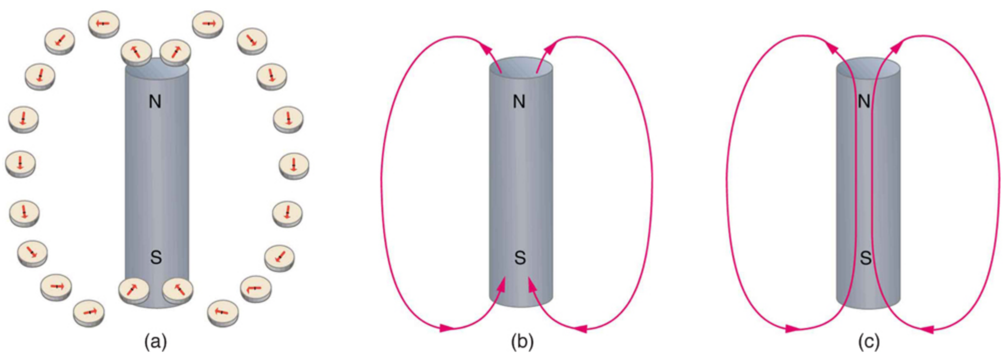
magnetic monopoles

an isolated magnetic pole; a south pole without a north pole, or vice versa (no magnetic monopole has ever been observed)

Magnetic Fields and Magnetic Field Lines

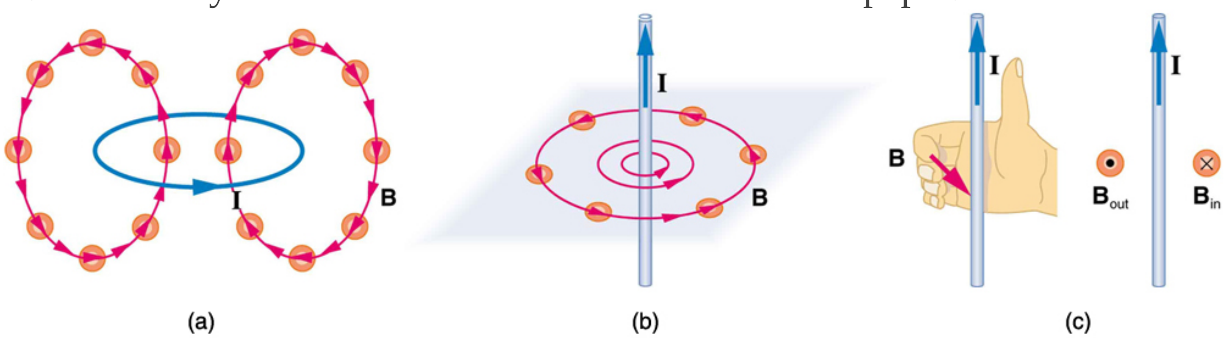
- Define magnetic field and describe the magnetic field lines of various magnetic fields.

Einstein is said to have been fascinated by a compass as a child, perhaps musing on how the needle felt a force without direct physical contact. His ability to think deeply and clearly about action at a distance, particularly for gravitational, electric, and magnetic forces, later enabled him to create his revolutionary theory of relativity. Since magnetic forces act at a distance, we define a **magnetic field** to represent magnetic forces. The pictorial representation of **magnetic field lines** is very useful in visualizing the strength and direction of the magnetic field. As shown in [link], the **direction of magnetic field lines** is defined to be the direction in which the north end of a compass needle points. The magnetic field is traditionally called the ***B*-field**.



Magnetic field lines are defined to have the direction that a small compass points when placed at a location. (a) If small compasses are used to map the magnetic field around a bar magnet, they will point in the directions shown: away from the north pole of the magnet, toward the south pole of the magnet. (Recall that the Earth's north magnetic pole is really a south pole in terms of definitions of poles on a bar magnet.) (b) Connecting the arrows gives continuous magnetic field lines. The strength of the field is proportional to the closeness (or density) of the lines. (c) If the interior of the magnet could be probed, the field lines would be found to form continuous closed loops.

Small compasses used to test a magnetic field will not disturb it. (This is analogous to the way we tested electric fields with a small test charge. In both cases, the fields represent only the object creating them and not the probe testing them.) [link] shows how the magnetic field appears for a current loop and a long straight wire, as could be explored with small compasses. A small compass placed in these fields will align itself parallel to the field line at its location, with its north pole pointing in the direction of *B*. Note the symbols used for field into and out of the paper.



Small compasses could be used to map the fields shown here. (a) The magnetic field of a circular current loop is similar to that of a bar magnet. (b) A long and straight wire creates a field with magnetic field lines forming circular loops. (c) When the wire is in the plane of the paper, the field is perpendicular to the paper. Note that the symbols used for the field pointing inward (like the tail of an arrow) and the field pointing outward (like the tip of an arrow).

Note:
Making Connections: Concept of a Field
A field is a way of mapping forces surrounding any object that can act on another object at a distance without apparent physical connection. The field represents the object generating it. Gravitational fields map

gravitational forces, electric fields map electrical forces, and magnetic fields map magnetic forces.

Extensive exploration of magnetic fields has revealed a number of hard-and-fast rules. We use magnetic field lines to represent the field (the lines are a pictorial tool, not a physical entity in and of themselves). The properties of magnetic field lines can be summarized by these rules:

1. The direction of the magnetic field is tangent to the field line at any point in space. A small compass will point in the direction of the field line.
2. The strength of the field is proportional to the closeness of the lines. It is exactly proportional to the number of lines per unit area perpendicular to the lines (called the areal density).
3. Magnetic field lines can never cross, meaning that the field is unique at any point in space.
4. Magnetic field lines are continuous, forming closed loops without beginning or end. They go from the north pole to the south pole.

The last property is related to the fact that the north and south poles cannot be separated. It is a distinct difference from electric field lines, which begin and end on the positive and negative charges. If magnetic monopoles existed, then magnetic field lines would begin and end on them.

Section Summary

- Magnetic fields can be pictorially represented by magnetic field lines, the properties of which are as follows:

1. The field is tangent to the magnetic field line.
2. Field strength is proportional to the line density.
3. Field lines cannot cross.
4. Field lines are continuous loops.

Conceptual Questions

Exercise:

Problem:

Explain why the magnetic field would not be unique (that is, not have a single value) at a point in space where magnetic field lines might cross. (Consider the direction of the field at such a point.)

Exercise:

Problem:

List the ways in which magnetic field lines and electric field lines are similar. For example, the field direction is tangent to the line at any point in space. Also list the ways in which they differ. For example, electric force is parallel to electric field lines, whereas magnetic force on moving charges is perpendicular to magnetic field lines.

Exercise:

Problem:

Noting that the magnetic field lines of a bar magnet resemble the electric field lines of a pair of equal and opposite charges, do you expect the magnetic field to rapidly decrease in strength with distance from the magnet? Is this consistent with your experience with magnets?

Exercise:

Problem:

Is the Earth’s magnetic field parallel to the ground at all locations? If not, where is it parallel to the surface? Is its strength the same at all locations? If not, where is it greatest?

Glossary

magnetic field
: the representation of magnetic forces

B-field
another term for magnetic field

magnetic field lines
the pictorial representation of the strength and the direction of a magnetic field

direction of magnetic field lines
the direction that the north end of a compass needle points

Magnetic Field Strength: Force on a Moving Charge in a Magnetic Field

- Describe the effects of magnetic fields on moving charges.
- Use the right hand rule 1 to determine the velocity of a charge, the direction of the magnetic field, and the direction of the magnetic force on a moving charge.
- Calculate the magnetic force on a moving charge.

What is the mechanism by which one magnet exerts a force on another? The answer is related to the fact that all magnetism is caused by current, the flow of charge. *Magnetic fields exert forces on moving charges*, and so they exert forces on other magnets, all of which have moving charges.

Right Hand Rule 1

The magnetic force on a moving charge is one of the most fundamental known. Magnetic force is as important as the electrostatic or Coulomb force. Yet the magnetic force is more complex, in both the number of factors that affects it and in its direction, than the relatively simple Coulomb force. The magnitude of the **magnetic force** F on a charge q moving at a speed v in a magnetic field of strength B is given by
Equation:

$$F = \text{qvB}\sin\theta,$$

where θ is the angle between the directions of $\mathbf{v}$ and $\mathbf{B}$. This force is often called the **Lorentz force**. In fact, this is how we define the magnetic field strength B—in terms of the force on a charged particle moving in a magnetic field. The SI unit for magnetic field strength B is called the **tesla** (T) after the eccentric but brilliant inventor Nikola Tesla (1856–1943). To determine how the tesla relates to other SI units, we solve $F = \text{qvB}\sin\theta$ for B.
Equation:

$$B = \frac{F}{\text{qv}\sin\theta}$$

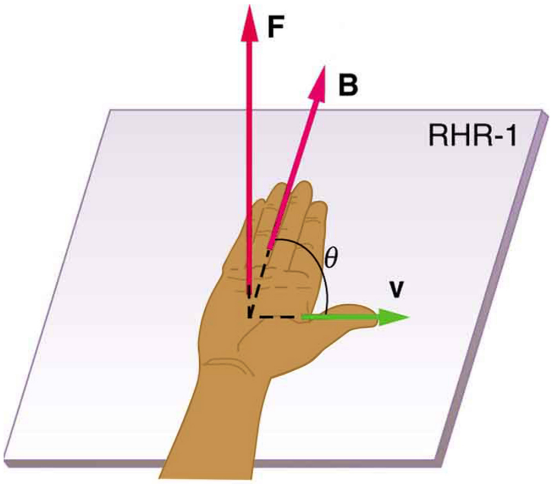
Because sin θ is unitless, the tesla is
Equation:

$$1\ \mathrm{T} = \frac{1\ \mathrm{N}}{\mathrm{C} \cdot \mathrm{m/s}} = \frac{1\ \mathrm{N}}{\mathrm{A} \cdot \mathrm{m}}$$

(note that C/s = A).

Another smaller unit, called the **gauss** (G), where $1\ \mathrm{G} = 10^{-4}\ \mathrm{T}$, is sometimes used. The strongest permanent magnets have fields near 2 T; superconducting electromagnets may attain 10 T or more. The Earth's magnetic field on its surface is only about $5 \times 10^{-5}\ \mathrm{T}$, or 0.5 G.

The *direction* of the magnetic force $\mathbf{F}$ is perpendicular to the plane formed by $\mathbf{v}$ and $\mathbf{B}$, as determined by the **right hand rule 1** (or RHR-1), which is illustrated in [link]. RHR-1 states that, to determine the direction of the magnetic force on a positive moving charge, you point the thumb of the right hand in the direction of $\mathbf{v}$, the fingers in the direction of $\mathbf{B}$, and a perpendicular to the palm points in the direction of $\mathbf{F}$. One way to remember this is that there is one velocity, and so the thumb represents it. There are many field lines, and so the fingers represent them. The force is in the direction you would push with your palm. The force on a negative charge is in exactly the opposite direction to that on a positive charge.



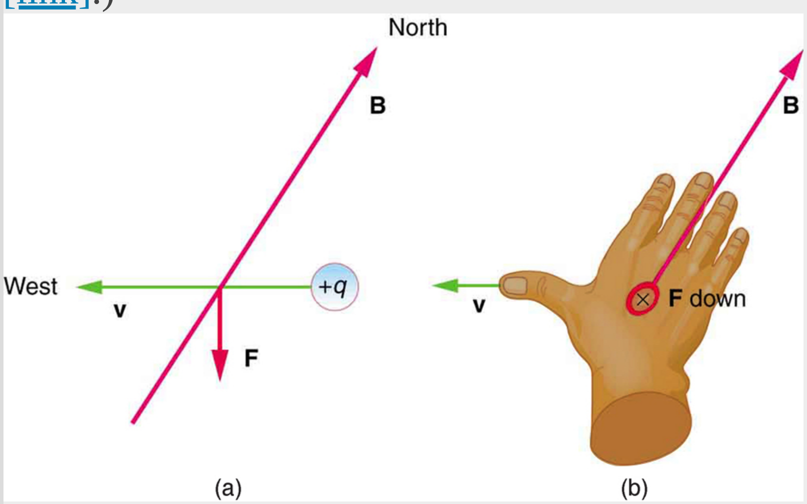
Magnetic fields exert forces on moving charges. This force is one of the most basic known. The direction of the magnetic force on a moving charge is perpendicular to the plane formed by $\mathbf{v}$ and $\mathbf{B}$ and follows right hand rule–1 (RHR-1) as shown. The magnitude of the force is proportional to q, v, B, and the sine of the angle between $\mathbf{v}$ and $\mathbf{B}$.

Note:
Making Connections: Charges and Magnets
There is no magnetic force on static charges. However, there is a magnetic force on moving charges. When charges are stationary, their electric fields do not affect magnets. But, when charges move, they produce magnetic

fields that exert forces on other magnets. When there is relative motion, a connection between electric and magnetic fields emerges—each affects the other.

Example:
Calculating Magnetic Force: Earth's Magnetic Field on a Charged Glass Rod

With the exception of compasses, you seldom see or personally experience forces due to the Earth's small magnetic field. To illustrate this, suppose that in a physics lab you rub a glass rod with silk, placing a 20-nC positive charge on it. Calculate the force on the rod due to the Earth's magnetic field, if you throw it with a horizontal velocity of 10 m/s due west in a place where the Earth's field is due north parallel to the ground. (The direction of the force is determined with right hand rule 1 as shown in [link].)



A positively charged object moving due west in a region where the Earth's magnetic field is due north experiences a force that is straight down as shown. A negative charge moving in the same direction would feel a force straight up.

Strategy
We are given the charge, its velocity, and the magnetic field strength and direction. We can thus use the equation $F = \text{qvB sin } \theta$ to find the force.
Solution
The magnetic force is
Equation:

$$F = \text{qvb sin } \theta.$$

We see that $\text{sin } \theta = 1$, since the angle between the velocity and the direction of the field is 90º. Entering the other given quantities yields
Equation:

$$\begin{array}{lll} F & = & \left(20 \times 10^{-9}\text{ C}\right)\left(10\text{ m/s}\right)\left(5 \times 10^{-5}\text{ T}\right) \\ & = & 1 \times 10^{-11}\text{ }\left(\text{C} \cdot \text{m/s}\right)\left(\frac{\text{N}}{\text{C} \cdot \text{m/s}}\right) = 1 \times 10^{-11}\text{ N.} \end{array}$$

Discussion
This force is completely negligible on any macroscopic object, consistent with experience. (It is calculated to only one digit, since the Earth's field varies with location and is given to only one digit.) The Earth's magnetic field, however, does produce very important effects, particularly on submicroscopic particles. Some of these are explored in Force on a Moving Charge in a Magnetic Field: Examples and Applications.

Section Summary

- Magnetic fields exert a force on a moving charge q, the magnitude of which is
  Equation:

  $$F = \text{qvB sin } \theta,$$

  where θ is the angle between the directions of v and B.
- The SI unit for magnetic field strength B is the tesla (T), which is related to other units by

Equation:

$$1\ \text{T} = \frac{1\ \text{N}}{\text{C} \cdot \text{m/s}} = \frac{1\ \text{N}}{\text{A} \cdot \text{m}}.$$

- The *direction* of the force on a moving charge is given by right hand rule 1 (RHR-1): Point the thumb of the right hand in the direction of v, the fingers in the direction of B, and a perpendicular to the palm points in the direction of F.
- The force is perpendicular to the plane formed by $\mathbf{v}$ and $\mathbf{B}$. Since the force is zero if $\mathbf{v}$ is parallel to $\mathbf{B}$, charged particles often follow magnetic field lines rather than cross them.

Conceptual Questions

Exercise:

Problem:

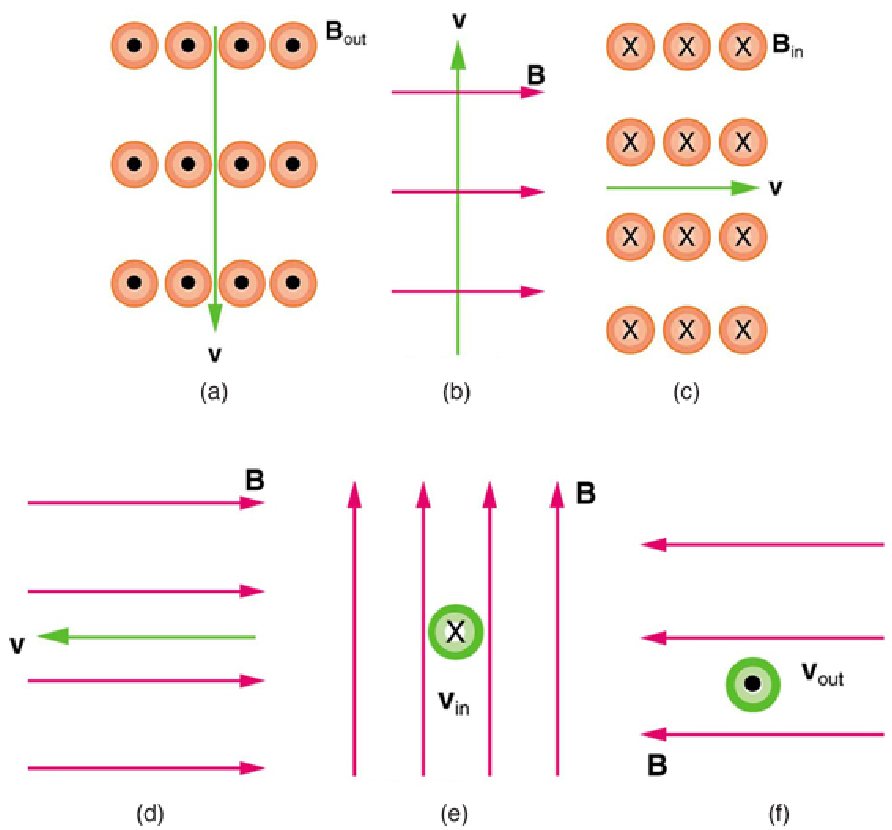
If a charged particle moves in a straight line through some region of space, can you say that the magnetic field in that region is necessarily zero?

Problems & Exercises

Exercise:

Problem:

What is the direction of the magnetic force on a positive charge that moves as shown in each of the six cases shown in [link]?



Solution:

(a) Left (West)

(b) Into the page

(c) Up (North)

(d) No force

(e) Right (East)

(f) Down (South)

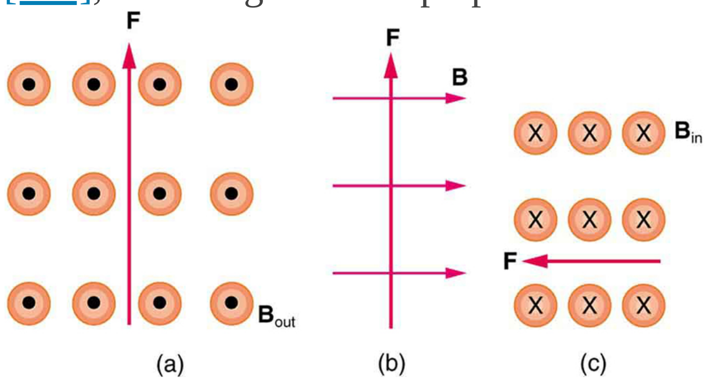
Exercise:

Problem: Repeat [link] for a negative charge.

Exercise:

Problem:

What is the direction of the velocity of a negative charge that experiences the magnetic force shown in each of the three cases in [link], assuming it moves perpendicular to $\mathbf{B}$?



Solution:

(a) East (right)

(b) Into page

(c) South (down)

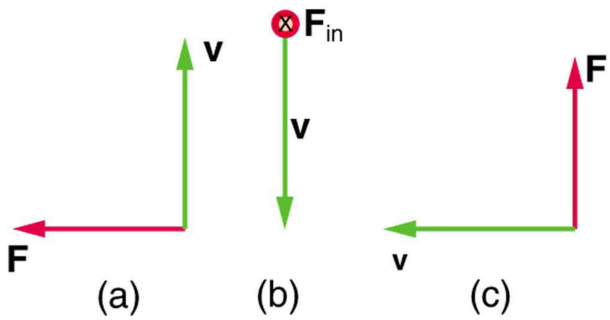
Exercise:

Problem: Repeat [link] for a positive charge.

Exercise:

Problem:

What is the direction of the magnetic field that produces the magnetic force on a positive charge as shown in each of the three cases in the figure below, assuming $\mathbf{B}$ is perpendicular to $\mathbf{v}$?



Solution:

(a) Into page

(b) West (left)

(c) Out of page

Exercise:

Problem: Repeat [link] for a negative charge.

Exercise:

Problem:

What is the maximum magnitude of the force on an aluminum rod with a $0.100\text{-}\mu\text{C}$ charge that you pass between the poles of a 1.50-T permanent magnet at a speed of 5.00 m/s? In what direction is the force?

Solution:

7.50×10^{-7} N perpendicular to both the magnetic field lines and the velocity

Exercise:

Problem:

(a) Aircraft sometimes acquire small static charges. Suppose a supersonic jet has a 0.500-μC charge and flies due west at a speed of 660 m/s over the Earth's magnetic south pole (near Earth's geographic north pole), where the $8.00 \times 10^{-5}\text{-T}$ magnetic field points straight down. What are the direction and the magnitude of the magnetic force on the plane? (b) Discuss whether the value obtained in part (a) implies this is a significant or negligible effect.

Exercise:

Problem:

(a) A cosmic ray proton moving toward the Earth at $5.00 \times 10^{7}\text{ m/s}$ experiences a magnetic force of $1.70 \times 10^{-16}\text{ N}$. What is the strength of the magnetic field if there is a $45º$ angle between it and the proton's velocity? (b) Is the value obtained in part (a) consistent with the known strength of the Earth's magnetic field on its surface? Discuss.

Solution:

(a) $3.01 \times 10^{-5}\text{ T}$

(b) This is slightly less then the magnetic field strength of $5 \times 10^{-5}\text{ T}$ at the surface of the Earth, so it is consistent.

Exercise:

Problem:

An electron moving at $4.00 \times 10^{3}\text{ m/s}$ in a 1.25-T magnetic field experiences a magnetic force of $1.40 \times 10^{-16}\text{ N}$. What angle does the velocity of the electron make with the magnetic field? There are two answers.

Exercise:

Problem:

(a) A physicist performing a sensitive measurement wants to limit the magnetic force on a moving charge in her equipment to less than 1.00×10^{-12} N. What is the greatest the charge can be if it moves at a maximum speed of 30.0 m/s in the Earth's field? (b) Discuss whether it would be difficult to limit the charge to less than the value found in (a) by comparing it with typical static electricity and noting that static is often absent.

Solution:

(a) 6.67×10^{-10} C (taking the Earth's field to be 5.00×10^{-5} T)

(b) Less than typical static, therefore difficult

Glossary

right hand rule 1 (RHR-1)
: the rule to determine the direction of the magnetic force on a positive moving charge: when the thumb of the right hand points in the direction of the charge's velocity **v** and the fingers point in the direction of the magnetic field $\mathbf{B}$, then the force on the charge is perpendicular and away from the palm; the force on a negative charge is perpendicular and into the palm

Lorentz force
: the force on a charge moving in a magnetic field

tesla
: T, the SI unit of the magnetic field strength; $1\ \text{T} = \frac{1\ \text{N}}{\text{A}\cdot\text{m}}$

magnetic force
: the force on a charge produced by its motion through a magnetic field; the Lorentz force

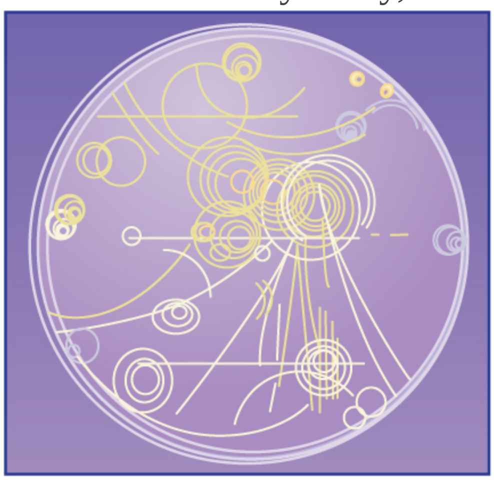
gauss

G, the unit of the magnetic field strength; $1\ \mathrm{G} = 10^{-4}\ \mathrm{T}$

Force on a Moving Charge in a Magnetic Field: Examples and Applications

- Describe the effects of a magnetic field on a moving charge.
- Calculate the radius of curvature of the path of a charge that is moving in a magnetic field.

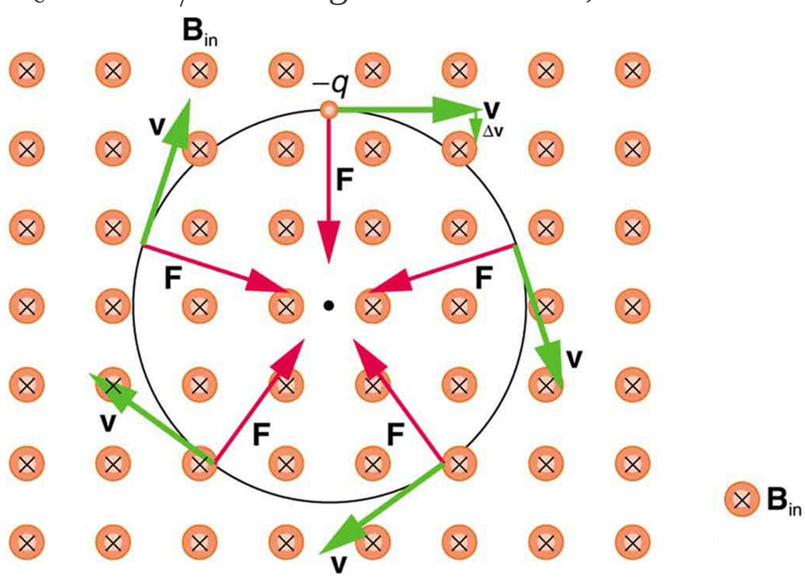
Magnetic force can cause a charged particle to move in a circular or spiral path. Cosmic rays are energetic charged particles in outer space, some of which approach the Earth. They can be forced into spiral paths by the Earth's magnetic field. Protons in giant accelerators are kept in a circular path by magnetic force. The bubble chamber photograph in [link] shows charged particles moving in such curved paths. The curved paths of charged particles in magnetic fields are the basis of a number of phenomena and can even be used analytically, such as in a mass spectrometer.



Trails of bubbles are produced by high-energy charged particles moving through the superheated liquid hydrogen in this artist's rendition of a bubble chamber. There is a strong magnetic field perpendicular to the page that causes the curved paths of the particles. The radius of the path can be

used to find the mass, charge, and energy of the particle.

So does the magnetic force cause circular motion? Magnetic force is always perpendicular to velocity, so that it does no work on the charged particle. The particle’s kinetic energy and speed thus remain constant. The direction of motion is affected, but not the speed. This is typical of uniform circular motion. The simplest case occurs when a charged particle moves perpendicular to a uniform B-field, such as shown in [link]. (If this takes place in a vacuum, the magnetic field is the dominant factor determining the motion.) Here, the magnetic force supplies the centripetal force $F_{\mathrm{c}} = mv^2/r$. Noting that $\sin\theta = 1$, we see that $F = \mathrm{qvB}$.



A negatively charged particle moves in the plane of the page in a region where the magnetic field is perpendicular into the page (represented by the small circles with x’s—like the tails of arrows). The magnetic force is perpendicular to the velocity, and so velocity changes in direction but not magnitude. Uniform circular motion results.

Because the magnetic force F supplies the centripetal force F_{c}, we have
Equation:

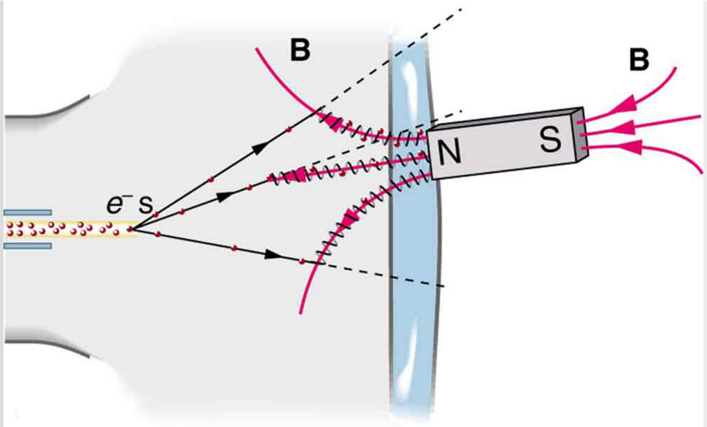
$$\text{qvB} = \frac{mv^2}{r}.$$

Solving for r yields
Equation:

$$r = \frac{\text{mv}}{\text{qB}}.$$

Here, r is the radius of curvature of the path of a charged particle with mass m and charge q, moving at a speed v perpendicular to a magnetic field of strength B. If the velocity is not perpendicular to the magnetic field, then v is the component of the velocity perpendicular to the field. The component of the velocity parallel to the field is unaffected, since the magnetic force is zero for motion parallel to the field. This produces a spiral motion rather than a circular one.

Example:
Calculating the Curvature of the Path of an Electron Moving in a Magnetic Field: A Magnet on a TV Screen
A magnet brought near an old-fashioned TV screen such as in [link] (TV sets with cathode ray tubes instead of LCD screens) severely distorts its picture by altering the path of the electrons that make its phosphors glow. ***(Don't try this at home, as it will permanently magnetize and ruin the TV.)*** To illustrate this, calculate the radius of curvature of the path of an electron having a velocity of $6.00 \times 10^7\ \text{m/s}$ (corresponding to the accelerating voltage of about 10.0 kV used in some TVs) perpendicular to a magnetic field of strength $B = 0.500\ \text{T}$ (obtainable with permanent magnets).



Side view showing what happens when a magnet comes in contact with a computer monitor or TV screen. Electrons moving toward the screen spiral about magnetic field lines, maintaining the component of their velocity parallel to the field lines. This distorts the image on the screen.

Strategy
We can find the radius of curvature r directly from the equation $r = \frac{mv}{qB}$, since all other quantities in it are given or known.
Solution
Using known values for the mass and charge of an electron, along with the given values of v and B gives us
Equation:

$$\begin{aligned} r = \frac{\text{mv}}{\text{qB}} &= \frac{\left(9.11\times10^{-31}\text{ kg}\right)\left(6.00\times10^{7}\text{ m/s}\right)}{\left(1.60\times10^{-19}\text{ C}\right)\left(0.500\text{ T}\right)} \\ &= 6.83\times10^{-4}\text{ m} \end{aligned}$$

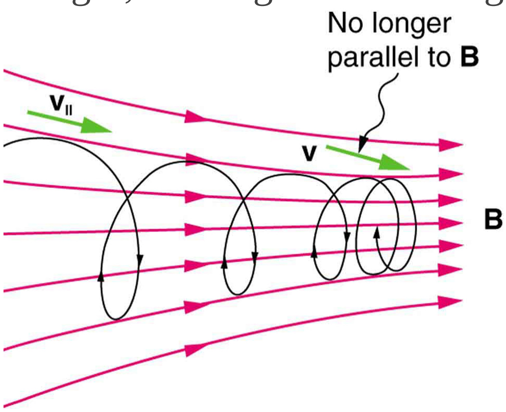
or
Equation:

$$r = 0.683\text{ mm.}$$

Discussion

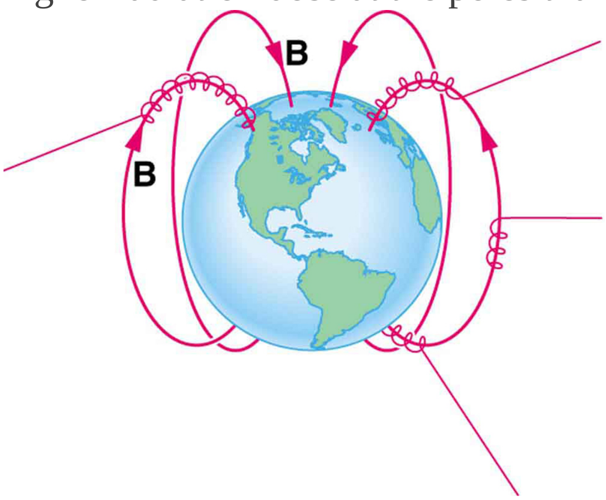
The small radius indicates a large effect. The electrons in the TV picture tube are made to move in very tight circles, greatly altering their paths and distorting the image.

[link] shows how electrons not moving perpendicular to magnetic field lines follow the field lines. The component of velocity parallel to the lines is unaffected, and so the charges spiral along the field lines. If field strength increases in the direction of motion, the field will exert a force to slow the charges, forming a kind of magnetic mirror, as shown below.



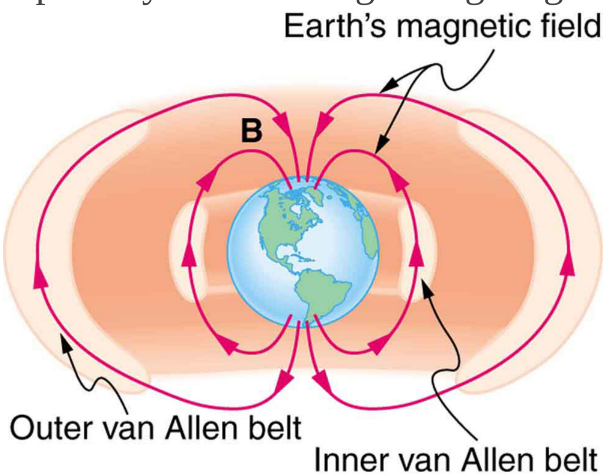
When a charged particle moves along a magnetic field line into a region where the field becomes stronger, the particle experiences a force that reduces the component of velocity parallel to the field. This force slows the motion along the field line and here reverses it, forming a “magnetic mirror.”

The properties of charged particles in magnetic fields are related to such different things as the Aurora Australis or Aurora Borealis and particle accelerators. *Charged particles approaching magnetic field lines may get trapped in spiral orbits about the lines rather than crossing them*, as seen above. Some cosmic rays, for example, follow the Earth's magnetic field lines, entering the atmosphere near the magnetic poles and causing the southern or northern lights through their ionization of molecules in the atmosphere. This glow of energized atoms and molecules is seen in [link]. Those particles that approach middle latitudes must cross magnetic field lines, and many are prevented from penetrating the atmosphere. Cosmic rays are a component of background radiation; consequently, they give a higher radiation dose at the poles than at the equator.



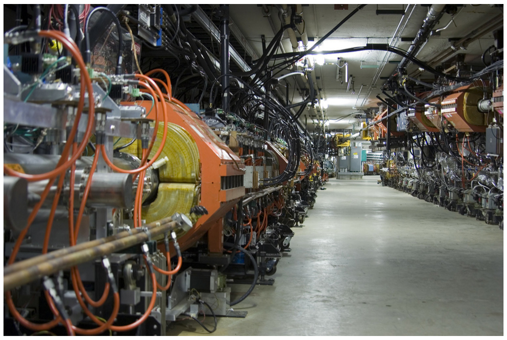
Energetic electrons and protons, components of cosmic rays, from the Sun and deep outer space often follow the Earth's magnetic field lines rather than cross them. (Recall that the Earth's north magnetic pole is really a south pole in terms of a bar magnet.)

Some incoming charged particles become trapped in the Earth’s magnetic field, forming two belts above the atmosphere known as the Van Allen radiation belts after the discoverer James A. Van Allen, an American astrophysicist. (See [link].) Particles trapped in these belts form radiation fields (similar to nuclear radiation) so intense that manned space flights avoid them and satellites with sensitive electronics are kept out of them. In the few minutes it took lunar missions to cross the Van Allen radiation belts, astronauts received radiation doses more than twice the allowed annual exposure for radiation workers. Other planets have similar belts, especially those having strong magnetic fields like Jupiter.



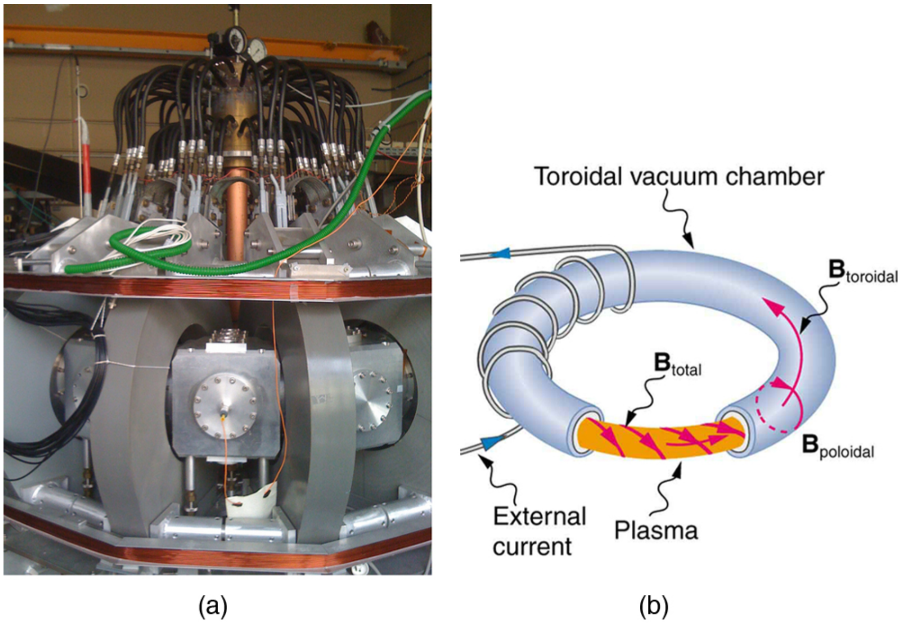
The Van Allen radiation belts are two regions in which energetic charged particles are trapped in the Earth’s magnetic field. One belt lies about 300 km above the Earth’s surface, the other about 16,000 km. Charged particles in these belts migrate along magnetic field lines and are partially reflected away from the poles by the stronger fields there. The charged particles that enter the atmosphere are replenished by the Sun and sources in deep outer space.

Back on Earth, we have devices that employ magnetic fields to contain charged particles. Among them are the giant particle accelerators that have been used to explore the substructure of matter. (See [link].) Magnetic fields not only control the direction of the charged particles, they also are used to focus particles into beams and overcome the repulsion of like charges in these beams.



The Fermilab facility in Illinois has a large particle accelerator (the most powerful in the world until 2008) that employs magnetic fields (magnets seen here in orange) to contain and direct its beam. This and other accelerators have been in use for several decades and have allowed us to discover some of the laws underlying all matter. (credit: ammcrim, Flickr)

Thermonuclear fusion (like that occurring in the Sun) is a hope for a future clean energy source. One of the most promising devices is the *tokamak*, which uses magnetic fields to contain (or trap) and direct the reactive charged particles. (See [link].) Less exotic, but more immediately practical, amplifiers in microwave ovens use a magnetic field to contain oscillating electrons. These oscillating electrons generate the microwaves sent into the oven.



Tokamaks such as the one shown in the figure are being studied with the goal of economical production of energy by nuclear fusion. Magnetic fields in the doughnut-shaped device contain and direct the reactive charged particles. (credit: David Mellis, Flickr)

Mass spectrometers have a variety of designs, and many use magnetic fields to measure mass. The curvature of a charged particle's path in the field is related to its mass and is measured to obtain mass information. (See More Applications of Magnetism.) Historically, such techniques were employed in the first direct observations of electron charge and mass. Today, mass

spectrometers (sometimes coupled with gas chromatographs) are used to determine the make-up and sequencing of large biological molecules.

Section Summary

- Magnetic force can supply centripetal force and cause a charged particle to move in a circular path of radius
  Equation:
  $$r=\frac{\text{mv}}{\text{qB}},$$
  where v is the component of the velocity perpendicular to B for a charged particle with mass m and charge q.

Conceptual Questions

Exercise:

Problem:

How can the motion of a charged particle be used to distinguish between a magnetic and an electric field?

Exercise:

Problem:

High-velocity charged particles can damage biological cells and are a component of radiation exposure in a variety of locations ranging from research facilities to natural background. Describe how you could use a magnetic field to shield yourself.

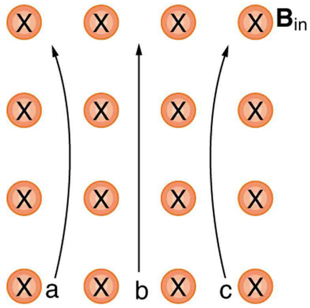
Exercise:

Problem:

If a cosmic ray proton approaches the Earth from outer space along a line toward the center of the Earth that lies in the plane of the equator, in what direction will it be deflected by the Earth's magnetic field? What about an electron? A neutron?

Exercise:

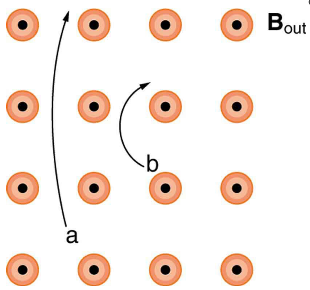
Problem: What are the signs of the charges on the particles in [link]?



Exercise:

Problem:

Which of the particles in [link] has the greatest velocity, assuming they have identical charges and masses?



Exercise:

Problem:

Which of the particles in [link] has the greatest mass, assuming all have identical charges and velocities?

Exercise:

Problem:

While operating, a high-precision TV monitor is placed on its side during maintenance. The image on the monitor changes color and blurs slightly. Discuss the possible relation of these effects to the Earth's magnetic field.

Problems & Exercises

If you need additional support for these problems, see More Applications of Magnetism.

Exercise:

Problem:

A cosmic ray electron moves at $7.50 \times 10^6\ \mathrm{m/s}$ perpendicular to the Earth's magnetic field at an altitude where field strength is $1.00 \times 10^{-5}\ \mathrm{T}$. What is the radius of the circular path the electron follows?

Solution:

4.27 m

Exercise:

Problem:

A proton moves at $7.50 \times 10^7\ \mathrm{m/s}$ perpendicular to a magnetic field. The field causes the proton to travel in a circular path of radius 0.800 m. What is the field strength?

Exercise:

Problem:

(a) Viewers of *Star Trek* hear of an antimatter drive on the Starship *Enterprise*. One possibility for such a futuristic energy source is to store antimatter charged particles in a vacuum chamber, circulating in a magnetic field, and then extract them as needed. Antimatter annihilates with normal matter, producing pure energy. What strength magnetic field is needed to hold antiprotons, moving at $5.00 \times 10^7\ \text{m/s}$ in a circular path 2.00 m in radius? Antiprotons have the same mass as protons but the opposite (negative) charge. (b) Is this field strength obtainable with today's technology or is it a futuristic possibility?

Solution:

(a) 0.261 T

(b) This strength is definitely obtainable with today's technology. Magnetic field strengths of 0.500 T are obtainable with permanent magnets.

Exercise:

Problem:

(a) An oxygen-16 ion with a mass of $2.66 \times 10^{-26}\ \text{kg}$ travels at $5.00 \times 10^6\ \text{m/s}$ perpendicular to a 1.20-T magnetic field, which makes it move in a circular arc with a 0.231-m radius. What positive charge is on the ion? (b) What is the ratio of this charge to the charge of an electron? (c) Discuss why the ratio found in (b) should be an integer.

Exercise:

Problem:

What radius circular path does an electron travel if it moves at the same speed and in the same magnetic field as the proton in [link]?

Solution:

4.36×10^{-4} m

Exercise:

Problem:

A velocity selector in a mass spectrometer uses a 0.100-T magnetic field. (a) What electric field strength is needed to select a speed of $4.00 \times 10^{6}\ \mathrm{m/s}$? (b) What is the voltage between the plates if they are separated by 1.00 cm?

Exercise:

Problem:

An electron in a TV CRT moves with a speed of $6.00 \times 10^{7}\ \mathrm{m/s}$, in a direction perpendicular to the Earth’s field, which has a strength of $5.00 \times 10^{-5}\ \mathrm{T}$. (a) What strength electric field must be applied perpendicular to the Earth’s field to make the electron moves in a straight line? (b) If this is done between plates separated by 1.00 cm, what is the voltage applied? (Note that TVs are usually surrounded by a ferromagnetic material to shield against external magnetic fields and avoid the need for such a correction.)

Solution:

(a) 3.00 kV/m

(b) 30.0 V

Exercise:

Problem:

(a) At what speed will a proton move in a circular path of the same radius as the electron in [link]? (b) What would the radius of the path be if the proton had the same speed as the electron? (c) What would the radius be if the proton had the same kinetic energy as the electron? (d) The same momentum?

Exercise:

Problem:

A mass spectrometer is being used to separate common oxygen-16 from the much rarer oxygen-18, taken from a sample of old glacial ice. (The relative abundance of these oxygen isotopes is related to climatic temperature at the time the ice was deposited.) The ratio of the masses of these two ions is 16 to 18, the mass of oxygen-16 is 2.66×10^{-26} kg, and they are singly charged and travel at 5.00×10^{6} m/s in a 1.20-T magnetic field. What is the separation between their paths when they hit a target after traversing a semicircle?

Solution:

0.173 m

Exercise:

Problem:

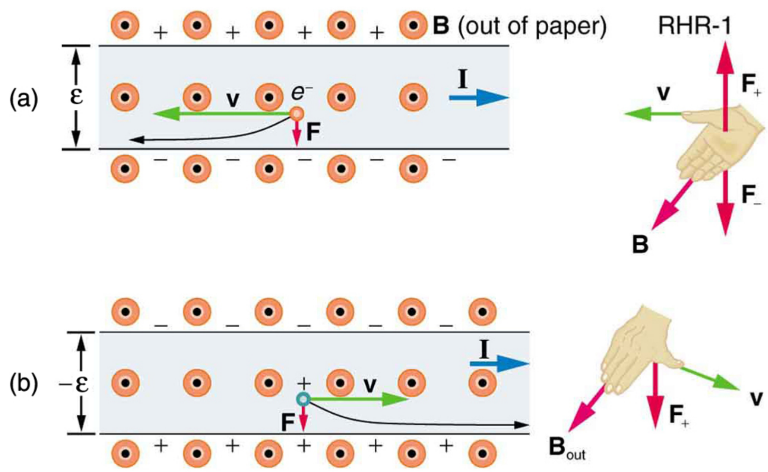
(a) Triply charged uranium-235 and uranium-238 ions are being separated in a mass spectrometer. (The much rarer uranium-235 is used as reactor fuel.) The masses of the ions are 3.90×10^{-25} kg and 3.95×10^{-25} kg, respectively, and they travel at 3.00×10^{5} m/s in a 0.250-T field. What is the separation between their paths when they hit a target after traversing a semicircle? (b) Discuss whether this distance between their paths seems to be big enough to be practical in the separation of uranium-235 from uranium-238.

The Hall Effect

- Describe the Hall effect.
- Calculate the Hall emf across a current-carrying conductor.

We have seen effects of a magnetic field on free-moving charges. The magnetic field also affects charges moving in a conductor. One result is the Hall effect, which has important implications and applications.

[link] shows what happens to charges moving through a conductor in a magnetic field. The field is perpendicular to the electron drift velocity and to the width of the conductor. Note that conventional current is to the right in both parts of the figure. In part (a), electrons carry the current and move to the left. In part (b), positive charges carry the current and move to the right. Moving electrons feel a magnetic force toward one side of the conductor, leaving a net positive charge on the other side. This separation of charge *creates a voltage* ε, known as the **Hall emf**, *across* the conductor. The creation of a voltage *across* a current-carrying conductor by a magnetic field is known as the **Hall effect**, after Edwin Hall, the American physicist who discovered it in 1879.



The Hall effect. (a) Electrons move to the left in this flat conductor (conventional current to the right). The magnetic field is directly out of the page, represented by circled dots; it exerts a force on the moving charges, causing a voltage ε, the

Hall emf, across the conductor. (b) Positive charges moving to the right (conventional current also to the right) are moved to the side, producing a Hall emf of the opposite sign, –ε. Thus, if the direction of the field and current are known, the sign of the charge carriers can be determined from the Hall effect.

One very important use of the Hall effect is to determine whether positive or negative charges carries the current. Note that in [link](b), where positive charges carry the current, the Hall emf has the sign opposite to when negative charges carry the current. Historically, the Hall effect was used to show that electrons carry current in metals and it also shows that positive charges carry current in some semiconductors. The Hall effect is used today as a research tool to probe the movement of charges, their drift velocities and densities, and so on, in materials. In 1980, it was discovered that the Hall effect is quantized, an example of quantum behavior in a macroscopic object.

The Hall effect has other uses that range from the determination of blood flow rate to precision measurement of magnetic field strength. To examine these quantitatively, we need an expression for the Hall emf, ε, across a conductor. Consider the balance of forces on a moving charge in a situation where B, v, and l are mutually perpendicular, such as shown in [link]. Although the magnetic force moves negative charges to one side, they cannot build up without limit. The electric field caused by their separation opposes the magnetic force, $F = \text{qvB}$, and the electric force, $F_{\text{e}} = \text{qE}$, eventually grows to equal it. That is,

Equation:

$$\text{qE} = \text{qvB}$$

or

Equation:

$$E = \text{vB}.$$

Note that the electric field E is uniform across the conductor because the magnetic field B is uniform, as is the conductor. For a uniform electric field, the relationship between electric field and voltage is $E = \varepsilon / l$, where l is the width of the conductor and ε is the Hall emf. Entering this into the last expression gives

Equation:

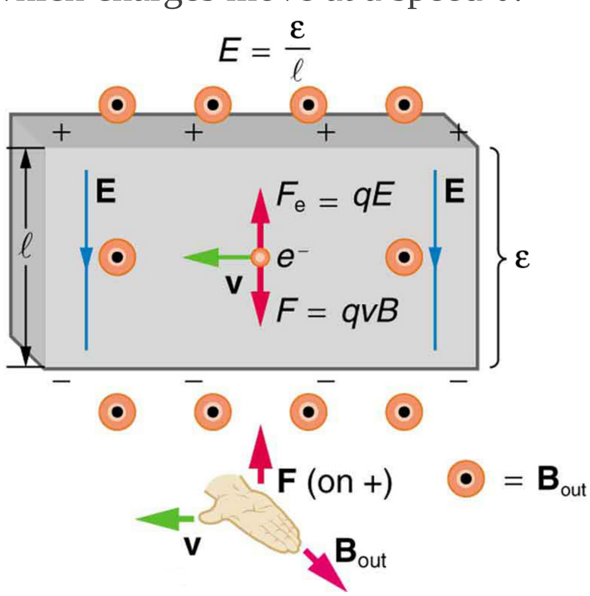
$$\frac{\varepsilon}{l} = \text{vB}.$$

Solving this for the Hall emf yields

Equation:

$$\varepsilon = Blv \; (B, v, \text{ and } l, \text{ mutually perpendicular}),$$

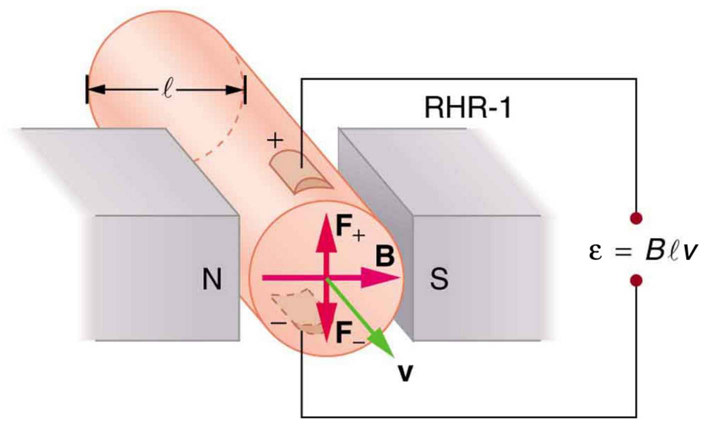
where ε is the Hall effect voltage across a conductor of width l through which charges move at a speed v.



The Hall emf ε produces an electric force that balances the magnetic force on the moving

charges. The magnetic force produces charge separation, which builds up until it is balanced by the electric force, an equilibrium that is quickly reached.

One of the most common uses of the Hall effect is in the measurement of magnetic field strength B. Such devices, called *Hall probes*, can be made very small, allowing fine position mapping. Hall probes can also be made very accurate, usually accomplished by careful calibration. Another application of the Hall effect is to measure fluid flow in any fluid that has free charges (most do). (See [link].) A magnetic field applied perpendicular to the flow direction produces a Hall emf ε as shown. Note that the sign of ε depends not on the sign of the charges, but only on the directions of B and v. The magnitude of the Hall emf is $\varepsilon = \text{Blv}$, where l is the pipe diameter, so that the average velocity v can be determined from ε providing the other factors are known.



The Hall effect can be used to measure fluid flow in any fluid having free charges, such as blood. The Hall emf ε is measured across the tube perpendicular to the applied magnetic field and is proportional to the average velocity v.

Example:
Calculating the Hall emf: Hall Effect for Blood Flow
A Hall effect flow probe is placed on an artery, applying a 0.100-T magnetic field across it, in a setup similar to that in [link]. What is the Hall emf, given the vessel's inside diameter is 4.00 mm and the average blood velocity is 20.0 cm/s?

Strategy
Because B, v, and l are mutually perpendicular, the equation $\varepsilon = \text{Blv}$ can be used to find ε.

Solution
Entering the given values for B, v, and l gives

Equation:

$$\begin{aligned} \varepsilon &= Blv = (0.100\text{ T})\ 4.00\times 10^{-3}\text{ m}\ (0.200\text{ m/s}) \\ &= 80.0\ \mu\text{V} \end{aligned}$$

Discussion
This is the average voltage output. Instantaneous voltage varies with pulsating blood flow. The voltage is small in this type of measurement. ε is particularly difficult to measure, because there are voltages associated with heart action (ECG voltages) that are on the order of millivolts. In practice, this difficulty is overcome by applying an AC magnetic field, so that the Hall emf is AC with the same frequency. An amplifier can be very selective in picking out only the appropriate frequency, eliminating signals and noise at other frequencies.

Section Summary

- The Hall effect is the creation of voltage ε, known as the Hall emf, across a current-carrying conductor by a magnetic field.
- The Hall emf is given by

Equation:

$$\varepsilon = Blv \; (B,\, v,\, \text{and } l,\, \text{mutually perpendicular})$$

for a conductor of width l through which charges move at a speed v.

Conceptual Questions

Exercise:

Problem:

Discuss how the Hall effect could be used to obtain information on free charge density in a conductor. (Hint: Consider how drift velocity and current are related.)

Problems & Exercises

Exercise:

Problem:

A large water main is 2.50 m in diameter and the average water velocity is 6.00 m/s. Find the Hall voltage produced if the pipe runs perpendicular to the Earth's $5.00 \times 10^{-5}\text{-T}$ field.

Solution:

$7.50 \times 10^{-4}\ \text{V}$

Exercise:

Problem:

What Hall voltage is produced by a 0.200-T field applied across a 2.60-cm-diameter aorta when blood velocity is 60.0 cm/s?

Exercise:

Problem:

(a) What is the speed of a supersonic aircraft with a 17.0-m wingspan, if it experiences a 1.60-V Hall voltage between its wing tips when in level flight over the north magnetic pole, where the Earth's field strength is $8.00 \times 10^{-5}\ \text{T}$? (b) Explain why very little current flows as a result of this Hall voltage.

Solution:

(a) 1.18×10^{3} m/s

(b) Once established, the Hall emf pushes charges one direction and the magnetic force acts in the opposite direction resulting in no net force on the charges. Therefore, no current flows in the direction of the Hall emf. This is the same as in a current-carrying conductor—current does not flow in the direction of the Hall emf.

Exercise:

Problem:

A nonmechanical water meter could utilize the Hall effect by applying a magnetic field across a metal pipe and measuring the Hall voltage produced. What is the average fluid velocity in a 3.00-cm-diameter pipe, if a 0.500-T field across it creates a 60.0-mV Hall voltage?

Exercise:

Problem:

Calculate the Hall voltage induced on a patient's heart while being scanned by an MRI unit. Approximate the conducting path on the heart wall by a wire 7.50 cm long that moves at 10.0 cm/s perpendicular to a 1.50-T magnetic field.

Solution:

11.3 mV

Exercise:

Problem:

A Hall probe calibrated to read 1.00 μV when placed in a 2.00-T field is placed in a 0.150-T field. What is its output voltage?

Exercise:

Problem:

Using information in [link], what would the Hall voltage be if a 2.00-T field is applied across a 10-gauge copper wire (2.588 mm in diameter) carrying a 20.0-A current?

Solution:

1.16 μV

Exercise:

Problem:

Show that the Hall voltage across wires made of the same material, carrying identical currents, and subjected to the same magnetic field is inversely proportional to their diameters. (Hint: Consider how drift velocity depends on wire diameter.)

Exercise:

Problem:

A patient with a pacemaker is mistakenly being scanned for an MRI image. A 10.0-cm-long section of pacemaker wire moves at a speed of 10.0 cm/s perpendicular to the MRI unit's magnetic field and a 20.0-mV Hall voltage is induced. What is the magnetic field strength?

Solution:

2.00 T

Glossary

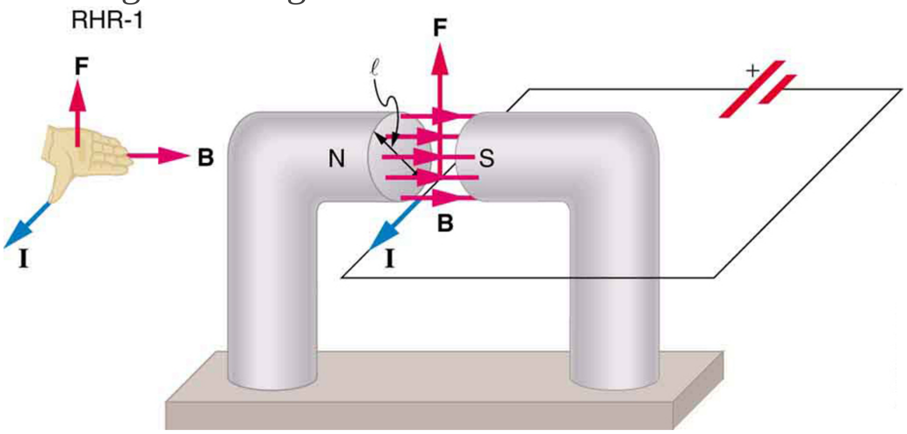
Hall effect
: the creation of voltage across a current-carrying conductor by a magnetic field

Hall emf
: the electromotive force created by a current-carrying conductor by a magnetic field, $\varepsilon = \mathrm{Blv}$

Magnetic Force on a Current-Carrying Conductor

- Describe the effects of a magnetic force on a current-carrying conductor.
- Calculate the magnetic force on a current-carrying conductor.

Because charges ordinarily cannot escape a conductor, the magnetic force on charges moving in a conductor is transmitted to the conductor itself.



The magnetic field exerts a force on a current-carrying wire in a direction given by the right hand rule 1 (the same direction as that on the individual moving charges). This force can easily be large enough to move the wire, since typical currents consist of very large numbers of moving charges.

We can derive an expression for the magnetic force on a current by taking a sum of the magnetic forces on individual charges. (The forces add because they are in the same direction.) The force on an individual charge moving at the drift velocity v_{d} is given by $F = qv_{\mathrm{d}}B \sin \theta$. Taking B to be uniform over a length of wire l and zero elsewhere, the total magnetic force on the wire is then $F = (qv_{\mathrm{d}}B \sin \theta)(N)$, where N is the number of charge carriers in the section of wire of length l. Now, $N = \mathrm{nV}$, where n is the number of charge carriers per unit volume and V is the volume of wire in the field. Noting that $V = \mathrm{Al}$, where A is the cross-sectional area of the

wire, then the force on the wire is $F = (qv_{\text{d}}B\sin\theta)(\text{nAl})$. Gathering terms,

Equation:

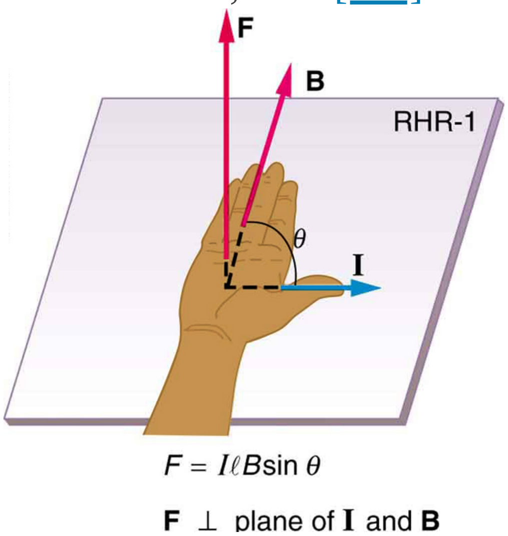
$$F = (nqAv_{\text{d}})lB\sin\theta.$$

Because $nqAv_{\text{d}} = I$ (see Current),

Equation:

$$F = \text{IlB}\sin\theta$$

is the equation for *magnetic force on a length* l *of wire carrying a current* I *in a uniform magnetic field* B, as shown in [link]. If we divide both sides of this expression by l, we find that the magnetic force per unit length of wire in a uniform field is $\frac{F}{l} = \text{IB}\sin\theta$. The direction of this force is given by RHR-1, with the thumb in the direction of the current I. Then, with the fingers in the direction of B, a perpendicular to the palm points in the direction of F, as in [link].



The force on a current-carrying wire in a magnetic field is $F = \text{IlB}\sin\theta$. Its

direction is given by
RHR-1.

Example:
Calculating Magnetic Force on a Current-Carrying Wire: A Strong Magnetic Field

Calculate the force on the wire shown in [link], given $B = 1.50\text{ T}$, $l = 5.00\text{ cm}$, and $I = 20.0\text{ A}$.

Strategy

The force can be found with the given information by using $F = \text{IlB sin }\theta$ and noting that the angle θ between I and B is 90º, so that $\text{sin }\theta = 1$.

Solution

Entering the given values into $F = \text{IlB sin }\theta$ yields

Equation:

$$F = \text{IlB sin }\theta = (20.0\text{ A})(0.0500\text{ m})(1.50\text{ T})(1).$$

The units for tesla are $1\text{ T} = \frac{\text{N}}{\text{A}\cdot\text{m}}$; thus,

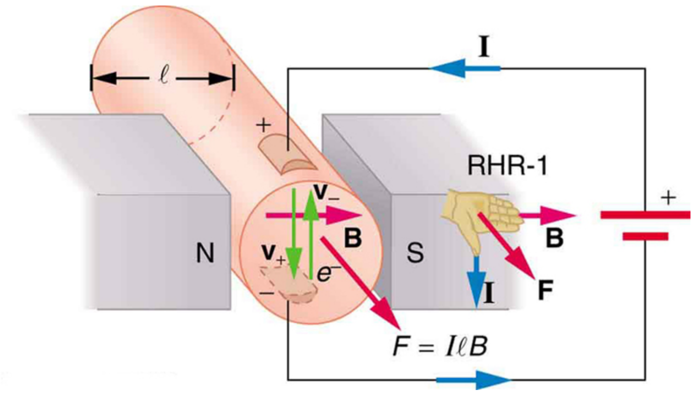
Equation:

$$F = 1.50\text{ N}.$$

Discussion

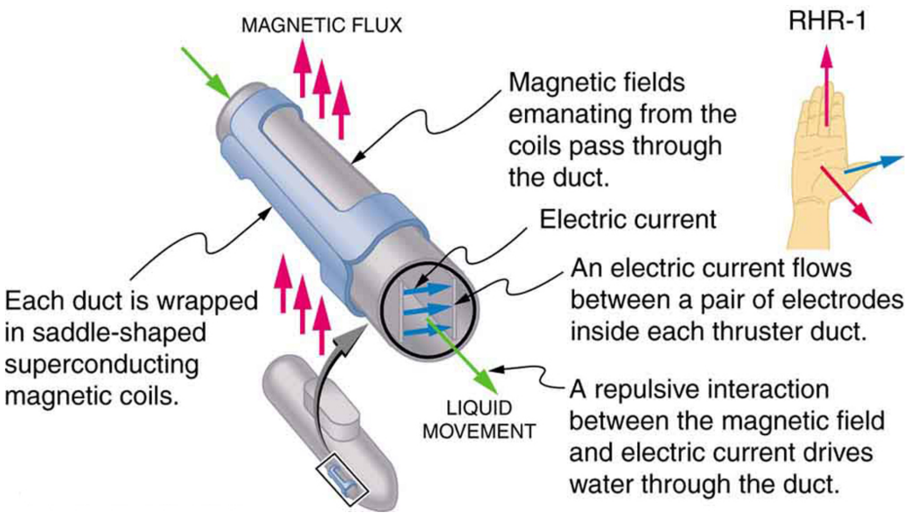
This large magnetic field creates a significant force on a small length of wire.

Magnetic force on current-carrying conductors is used to convert electric energy to work. (Motors are a prime example—they employ loops of wire and are considered in the next section.) Magnetohydrodynamics (MHD) is the technical name given to a clever application where magnetic force pumps fluids without moving mechanical parts. (See [link].)



Magnetohydrodynamics. The magnetic force on the current passed through this fluid can be used as a nonmechanical pump.

A strong magnetic field is applied across a tube and a current is passed through the fluid at right angles to the field, resulting in a force on the fluid parallel to the tube axis as shown. The absence of moving parts makes this attractive for moving a hot, chemically active substance, such as the liquid sodium employed in some nuclear reactors. Experimental artificial hearts are testing with this technique for pumping blood, perhaps circumventing the adverse effects of mechanical pumps. (Cell membranes, however, are affected by the large fields needed in MHD, delaying its practical application in humans.) MHD propulsion for nuclear submarines has been proposed, because it could be considerably quieter than conventional propeller drives. The deterrent value of nuclear submarines is based on their ability to hide and survive a first or second nuclear strike. As we slowly disassemble our nuclear weapons arsenals, the submarine branch will be the last to be decommissioned because of this ability (See [link].) Existing MHD drives are heavy and inefficient—much development work is needed.



An MHD propulsion system in a nuclear submarine could produce significantly less turbulence than propellers and allow it to run more silently. The development of a silent drive submarine was dramatized in the book and the film *The Hunt for Red October*.

Section Summary

- The magnetic force on current-carrying conductors is given by **Equation:**

  $$F = \text{IlB sin θ},$$

  where I is the current, l is the length of a straight conductor in a uniform magnetic field B, and θ is the angle between I and B. The force follows RHR-1 with the thumb in the direction of I.

Conceptual Questions

Exercise:

Problem:

Draw a sketch of the situation in [link] showing the direction of electrons carrying the current, and use RHR-1 to verify the direction of the force on the wire.

Exercise:

Problem:

Verify that the direction of the force in an MHD drive, such as that in [link], does not depend on the sign of the charges carrying the current across the fluid.

Exercise:

Problem:

Why would a magnetohydrodynamic drive work better in ocean water than in fresh water? Also, why would superconducting magnets be desirable?

Exercise:

Problem:

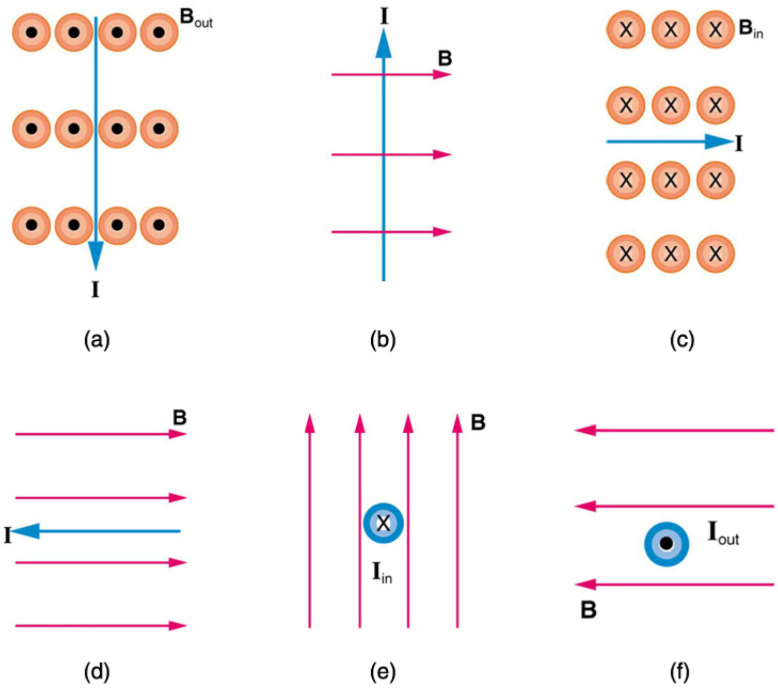
Which is more likely to interfere with compass readings, AC current in your refrigerator or DC current when you start your car? Explain.

Problems & Exercises

Exercise:

Problem:

What is the direction of the magnetic force on the current in each of the six cases in [link]?



Solution:

(a) west (left)

(b) into page

(c) north (up)

(d) no force

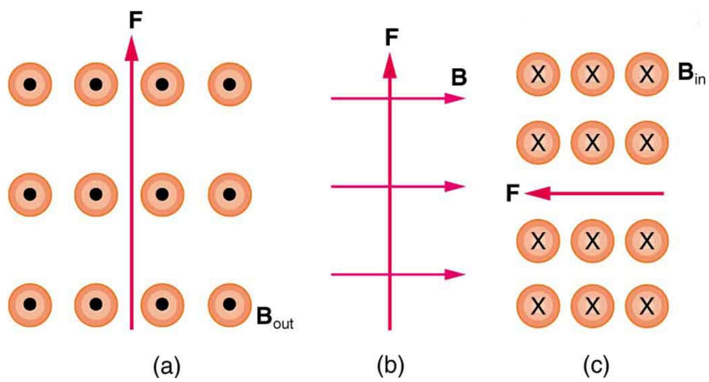
(e) east (right)

(f) south (down)

Exercise:

Problem:

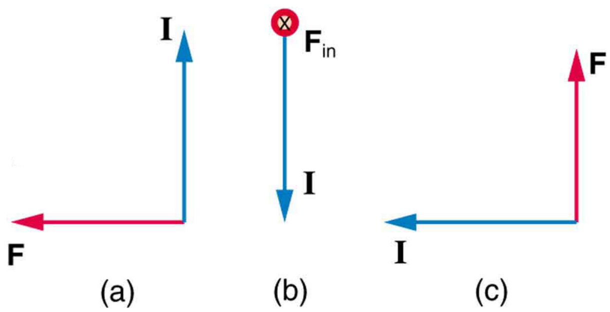
What is the direction of a current that experiences the magnetic force shown in each of the three cases in [link], assuming the current runs perpendicular to B?



Exercise:

Problem:

What is the direction of the magnetic field that produces the magnetic force shown on the currents in each of the three cases in [link], assuming is perpendicular to ?



Solution:

(a) into page

(b) west (left)

(c) out of page

Exercise:

Problem:

(a) What is the force per meter on a lightning bolt at the equator that carries 20,000 A perpendicular to the Earth's $3.00 \times 10^{-5}\text{-T}$ field? (b) What is the direction of the force if the current is straight up and the Earth's field direction is due north, parallel to the ground?

Exercise:

Problem:

(a) A DC power line for a light-rail system carries 1000 A at an angle of 30.0° to the Earth's 5.00×10^{-5} -T field. What is the force on a 100-m section of this line? (b) Discuss practical concerns this presents, if any.

Solution:

(a) 2.50 N

(b) This is about half a pound of force per 100 m of wire, which is much less than the weight of the wire itself. Therefore, it does not cause any special concerns.

Exercise:

Problem:

What force is exerted on the water in an MHD drive utilizing a 25.0-cm-diameter tube, if 100-A current is passed across the tube that is perpendicular to a 2.00-T magnetic field? (The relatively small size of this force indicates the need for very large currents and magnetic fields to make practical MHD drives.)

Exercise:

Problem:

A wire carrying a 30.0-A current passes between the poles of a strong magnet that is perpendicular to its field and experiences a 2.16-N force on the 4.00 cm of wire in the field. What is the average field strength?

Solution:

1.80 T

Exercise:

Problem:

(a) A 0.750-m-long section of cable carrying current to a car starter motor makes an angle of 60º with the Earth's $5.50 \times 10^{-5}\ \text{T}$ field. What is the current when the wire experiences a force of $7.00 \times 10^{-3}\ \text{N}$? (b) If you run the wire between the poles of a strong horseshoe magnet, subjecting 5.00 cm of it to a 1.75-T field, what force is exerted on this segment of wire?

Exercise:

Problem:

(a) What is the angle between a wire carrying an 8.00-A current and the 1.20-T field it is in if 50.0 cm of the wire experiences a magnetic force of 2.40 N? (b) What is the force on the wire if it is rotated to make an angle of 90º with the field?

Solution:

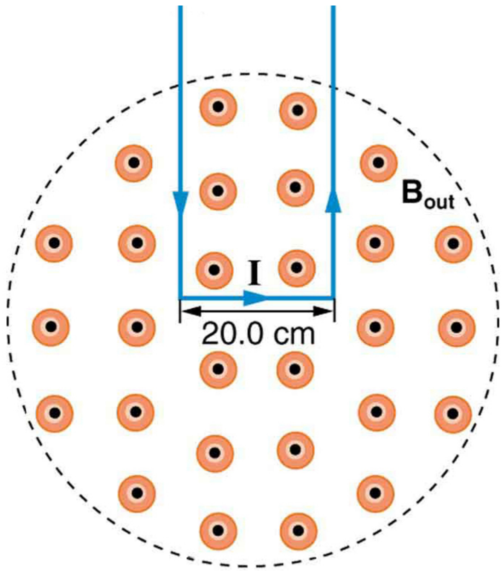
(a) 30º

(b) 4.80 N

Exercise:

Problem:

The force on the rectangular loop of wire in the magnetic field in [link] can be used to measure field strength. The field is uniform, and the plane of the loop is perpendicular to the field. (a) What is the direction of the magnetic force on the loop? Justify the claim that the forces on the sides of the loop are equal and opposite, independent of how much of the loop is in the field and do not affect the net force on the loop. (b) If a current of 5.00 A is used, what is the force per tesla on the 20.0-cm-wide loop?

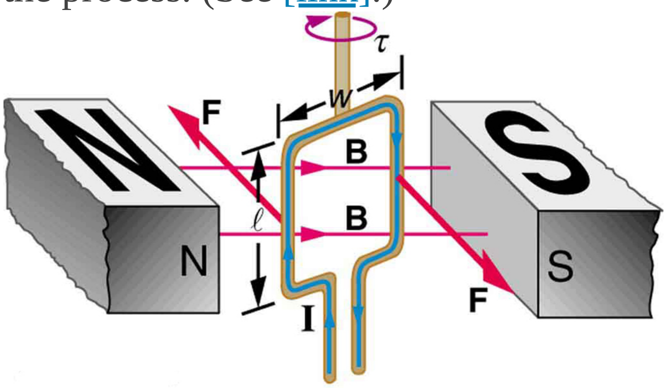


A rectangular loop of wire carrying a current is perpendicular to a magnetic field. The field is uniform in the region shown and is zero outside that region.

Torque on a Current Loop: Motors and Meters

- Describe how motors and meters work in terms of torque on a current loop.
- Calculate the torque on a current-carrying loop in a magnetic field.

Motors are the most common application of magnetic force on current-carrying wires. Motors have loops of wire in a magnetic field. When current is passed through the loops, the magnetic field exerts torque on the loops, which rotates a shaft. Electrical energy is converted to mechanical work in the process. (See [link].)



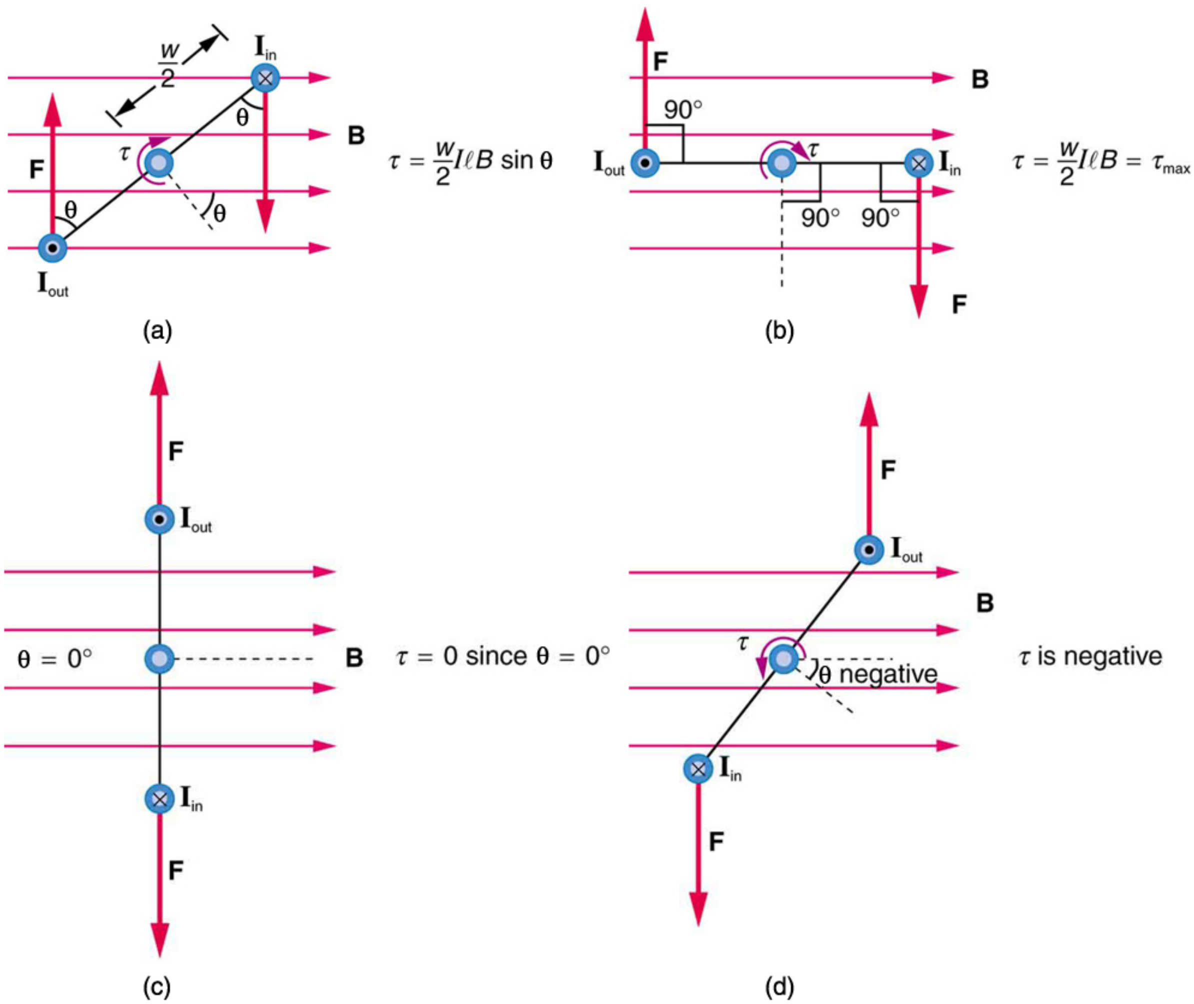
Torque on a current loop. A current-carrying loop of wire attached to a vertically rotating shaft feels magnetic forces that produce a clockwise torque as viewed from above.

Let us examine the force on each segment of the loop in [link] to find the torques produced about the axis of the vertical shaft. (This will lead to a useful equation for the torque on the loop.) We take the magnetic field to be uniform over the rectangular loop, which has width w and height l. First, we note that the forces on the top and bottom segments are vertical and, therefore, parallel to the shaft, producing no torque. Those vertical forces are equal in magnitude and opposite in direction, so that they also produce no net force on the loop. [link] shows views of the loop from above. Torque

is defined as $\tau = \mathrm{rF}\ \sin\ \theta$, where F is the force, r is the distance from the pivot that the force is applied, and θ is the angle between r and F. As seen in [link](a), right hand rule 1 gives the forces on the sides to be equal in magnitude and opposite in direction, so that the net force is again zero. However, each force produces a clockwise torque. Since $r = w/2$, the torque on each vertical segment is $(w/2)F\ \sin\ \theta$, and the two add to give a total torque.

Equation:

$$\tau = \frac{w}{2}F\ \sin\ \theta + \frac{w}{2}F\ \sin\ \theta = \mathrm{wF}\ \sin\ \theta$$



Top views of a current-carrying loop in a magnetic field. (a) The equation for torque is derived using this view. Note that the perpendicular to the loop makes an angle θ with the field that is the

same as the angle between $w/2$ and $\mathbf{F}$. (b) The maximum torque occurs when θ is a right angle and $\sin\theta = 1$. (c) Zero (minimum) torque occurs when θ is zero and $\sin\theta = 0$. (d) The torque reverses once the loop rotates past $\theta = 0$.

Now, each vertical segment has a length l that is perpendicular to B, so that the force on each is $F = \text{IlB}$. Entering F into the expression for torque yields

Equation:

$$\tau = \text{wIlB}\sin\theta.$$

If we have a multiple loop of N turns, we get N times the torque of one loop. Finally, note that the area of the loop is $A = \text{wl}$; the expression for the torque becomes

Equation:

$$\tau = \text{NIAB}\sin\theta.$$

This is the torque on a current-carrying loop in a uniform magnetic field. This equation can be shown to be valid for a loop of any shape. The loop carries a current I, has N turns, each of area A, and the perpendicular to the loop makes an angle θ with the field B. The net force on the loop is zero.

Example:

Calculating Torque on a Current-Carrying Loop in a Strong Magnetic Field

Find the maximum torque on a 100-turn square loop of a wire of 10.0 cm on a side that carries 15.0 A of current in a 2.00-T field.

Strategy

Torque on the loop can be found using $\tau = \text{NIAB}\sin\theta$. Maximum torque occurs when $\theta = 90^{\circ}$ and $\sin\theta = 1$.

Solution
For $\sin\theta = 1$, the maximum torque is
Equation:

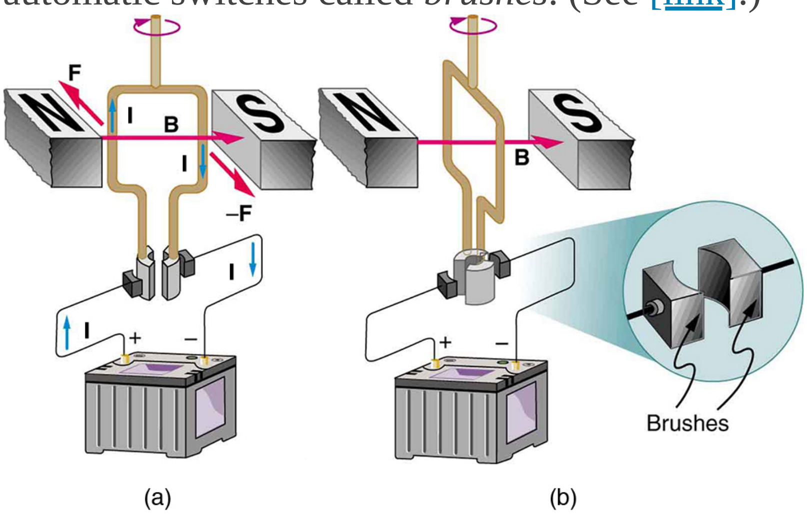
$$\tau_{\text{max}} = \text{NIAB}.$$

Entering known values yields
Equation:

$$\begin{aligned}\tau_{\text{max}} &= (100)(15.0\ \text{A})\left(0.100\ \text{m}^2\right)(2.00\ \text{T}) \\ &= 30.0\ \text{N}\cdot\text{m}.\end{aligned}$$

Discussion
This torque is large enough to be useful in a motor.

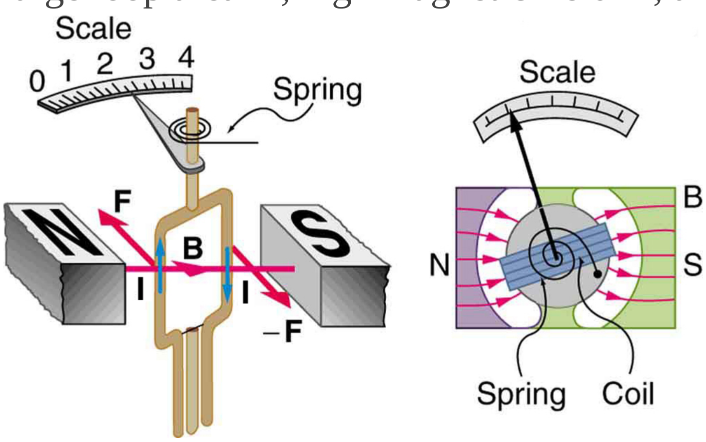
The torque found in the preceding example is the maximum. As the coil rotates, the torque decreases to zero at $\theta = 0$. The torque then *reverses* its direction once the coil rotates past $\theta = 0$. (See [link](d).) This means that, unless we do something, the coil will oscillate back and forth about equilibrium at $\theta = 0$. To get the coil to continue rotating in the same direction, we can reverse the current as it passes through $\theta = 0$ with automatic switches called *brushes*. (See [link].)



(a) As the angular momentum of the coil carries it through $\theta = 0$, the brushes reverse

the current to keep the torque clockwise. (b) The coil will rotate continuously in the clockwise direction, with the current reversing each half revolution to maintain the clockwise torque.

Meters, such as those in analog fuel gauges on a car, are another common application of magnetic torque on a current-carrying loop. [link] shows that a meter is very similar in construction to a motor. The meter in the figure has its magnets shaped to limit the effect of θ by making B perpendicular to the loop over a large angular range. Thus the torque is proportional to I and not θ. A linear spring exerts a counter-torque that balances the current-produced torque. This makes the needle deflection proportional to I. If an exact proportionality cannot be achieved, the gauge reading can be calibrated. To produce a galvanometer for use in analog voltmeters and ammeters that have a low resistance and respond to small currents, we use a large loop area A, high magnetic field B, and low-resistance coils.



Meters are very similar to motors but only rotate through a part of a revolution. The magnetic poles of this meter are shaped to keep the component of B perpendicular to the loop constant, so that the torque does not depend on θ and the deflection

against the return spring is proportional only to the current I.

Section Summary

- The torque τ on a current-carrying loop of any shape in a uniform magnetic field. is
  Equation:

  $$\tau = \text{NIAB}\sin\theta,$$

  where N is the number of turns, I is the current, A is the area of the loop, B is the magnetic field strength, and θ is the angle between the perpendicular to the loop and the magnetic field.

Conceptual Questions

Exercise:

Problem:

Draw a diagram and use RHR-1 to show that the forces on the top and bottom segments of the motor's current loop in [link] are vertical and produce no torque about the axis of rotation.

Problems & Exercises

Exercise:

Problem:

(a) By how many percent is the torque of a motor decreased if its permanent magnets lose 5.0% of their strength? (b) How many percent would the current need to be increased to return the torque to original values?

Solution:

(a) τ decreases by 5.00% if B decreases by 5.00%

(b) 5.26% increase

Exercise:

Problem:

(a) What is the maximum torque on a 150-turn square loop of wire 18.0 cm on a side that carries a 50.0-A current in a 1.60-T field? (b) What is the torque when θ is 10.9º?

Exercise:

Problem:

Find the current through a loop needed to create a maximum torque of $9.00\ \mathrm{N}\cdot\mathrm{m}$. The loop has 50 square turns that are 15.0 cm on a side and is in a uniform 0.800-T magnetic field.

Solution:

10.0 A

Exercise:

Problem:

Calculate the magnetic field strength needed on a 200-turn square loop 20.0 cm on a side to create a maximum torque of $300\ \mathrm{N}\cdot\mathrm{m}$ if the loop is carrying 25.0 A.

Exercise:

Problem:

Since the equation for torque on a current-carrying loop is $\tau = \mathrm{NIAB}\ \sin\ \theta$, the units of $\mathrm{N}\cdot\mathrm{m}$ must equal units of $\mathrm{A}\cdot\mathrm{m}^2\ \mathrm{T}$. Verify this.

Solution:

$$A \cdot m^2 \cdot T = A \cdot m^2 \left(\frac{N}{A \cdot m}\right) = N \cdot m.$$

Exercise:

Problem:

(a) At what angle θ is the torque on a current loop 90.0% of maximum? (b) 50.0% of maximum? (c) 10.0% of maximum?

Exercise:

Problem:

A proton has a magnetic field due to its spin on its axis. The field is similar to that created by a circular current loop $0.650 \times 10^{-15}\ \text{m}$ in radius with a current of $1.05 \times 10^4\ \text{A}$ (no kidding). Find the maximum torque on a proton in a 2.50-T field. (This is a significant torque on a small particle.)

Solution:

$3.48 \times 10^{-26}\ \text{N} \cdot \text{m}$

Exercise:

Problem:

(a) A 200-turn circular loop of radius 50.0 cm is vertical, with its axis on an east-west line. A current of 100 A circulates clockwise in the loop when viewed from the east. The Earth’s field here is due north, parallel to the ground, with a strength of $3.00 \times 10^{-5}\ \text{T}$. What are the direction and magnitude of the torque on the loop? (b) Does this device have any practical applications as a motor?

Exercise:

Problem:

Repeat [link], but with the loop lying flat on the ground with its current circulating counterclockwise (when viewed from above) in a location where the Earth’s field is north, but at an angle 45.0^{o} below the horizontal and with a strength of $6.00 \times 10^{-5}\ \text{T}$.

Solution:

(a) $0.666 \text{ N} \cdot \text{m}$ west

(b) This is not a very significant torque, so practical use would be limited. Also, the current would need to be alternated to make the loop rotate (otherwise it would oscillate).

Glossary

motor
: loop of wire in a magnetic field; when current is passed through the loops, the magnetic field exerts torque on the loops, which rotates a shaft; electrical energy is converted to mechanical work in the process

meter
: common application of magnetic torque on a current-carrying loop that is very similar in construction to a motor; by design, the torque is proportional to I and not θ, so the needle deflection is proportional to the current

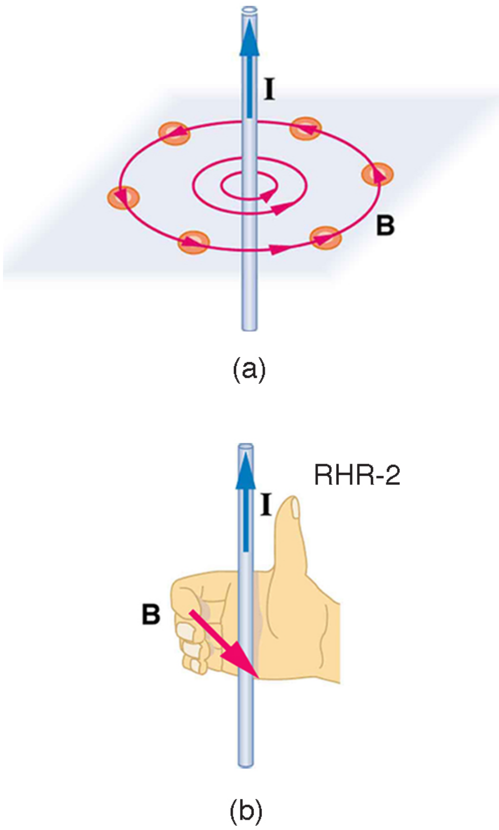
Magnetic Fields Produced by Currents: Ampere's Law

- Calculate current that produces a magnetic field.
- Use the right hand rule 2 to determine the direction of current or the direction of magnetic field loops.

How much current is needed to produce a significant magnetic field, perhaps as strong as the Earth's field? Surveyors will tell you that overhead electric power lines create magnetic fields that interfere with their compass readings. Indeed, when Oersted discovered in 1820 that a current in a wire affected a compass needle, he was not dealing with extremely large currents. How does the shape of wires carrying current affect the shape of the magnetic field created? We noted earlier that a current loop created a magnetic field similar to that of a bar magnet, but what about a straight wire or a toroid (doughnut)? How is the direction of a current-created field related to the direction of the current? Answers to these questions are explored in this section, together with a brief discussion of the law governing the fields created by currents.

Magnetic Field Created by a Long Straight Current-Carrying Wire: Right Hand Rule 2

Magnetic fields have both direction and magnitude. As noted before, one way to explore the direction of a magnetic field is with compasses, as shown for a long straight current-carrying wire in [link]. Hall probes can determine the magnitude of the field. The field around a long straight wire is found to be in circular loops. The **right hand rule 2** (RHR-2) emerges from this exploration and is valid for any current segment—*point the thumb in the direction of the current, and the fingers curl in the direction of the magnetic field loops* created by it.



(a) Compasses placed near a long straight current-carrying wire indicate that field lines form circular loops centered on the wire. (b) Right hand rule 2 states that, if the right hand thumb points in the direction of the current, the fingers curl in the direction of the field. This rule is consistent with the field mapped for the long straight wire and is valid for any current segment.

The **magnetic field strength (magnitude) produced by a long straight current-carrying wire** is found by experiment to be

Equation:

$$B=\frac{\mu_0 I}{2\pi r}\ (\text{long straight wire}),$$

where I is the current, r is the shortest distance to the wire, and the constant $\mu_0 = 4\pi\ \times\ 10^{-7}\ \text{T}\cdot\text{m/A}$ is the **permeability of free space**. (μ_0 is one of the basic constants in nature. We will see later that μ_0 is related to the speed of light.) Since the wire is very long, the magnitude of the field depends only on distance from the wire r, not on position along the wire.

Example:

Calculating Current that Produces a Magnetic Field

Find the current in a long straight wire that would produce a magnetic field twice the strength of the Earth's at a distance of 5.0 cm from the wire.

Strategy

The Earth's field is about $5.0\times10^{-5}\ \text{T}$, and so here B due to the wire is taken to be $1.0\times10^{-4}\ \text{T}$. The equation $B=\frac{\mu_0 I}{2\pi r}$ can be used to find I, since all other quantities are known.

Solution

Solving for I and entering known values gives

Equation:

$$\begin{aligned} I &= \frac{2\pi rB}{\mu_0} = \frac{2\pi\left(5.0\times10^{-2}\ \text{m}\right)\left(1.0\times10^{-4}\ \text{T}\right)}{4\pi\times10^{-7}\ \text{T}\cdot\text{m/A}} \\ &= 25\ \text{A}. \end{aligned}$$

Discussion

So a moderately large current produces a significant magnetic field at a distance of 5.0 cm from a long straight wire. Note that the answer is stated to only two digits, since the Earth's field is specified to only two digits in this example.

Ampere’s Law and Others

The magnetic field of a long straight wire has more implications than you might at first suspect. *Each segment of current produces a magnetic field like that of a long straight wire, and the total field of any shape current is the vector sum of the fields due to each segment.* The formal statement of the direction and magnitude of the field due to each segment is called the **Biot-Savart law**. Integral calculus is needed to sum the field for an arbitrary shape current. This results in a more complete law, called **Ampere’s law**, which relates magnetic field and current in a general way. Ampere’s law in turn is a part of **Maxwell’s equations**, which give a complete theory of all electromagnetic phenomena. Considerations of how Maxwell’s equations appear to different observers led to the modern theory of relativity, and the realization that electric and magnetic fields are different manifestations of the same thing. Most of this is beyond the scope of this text in both mathematical level, requiring calculus, and in the amount of space that can be devoted to it. But for the interested student, and particularly for those who continue in physics, engineering, or similar pursuits, delving into these matters further will reveal descriptions of nature that are elegant as well as profound. In this text, we shall keep the general features in mind, such as RHR-2 and the rules for magnetic field lines listed in Magnetic Fields and Magnetic Field Lines, while concentrating on the fields created in certain important situations.

Note:
Making Connections: Relativity
Hearing all we do about Einstein, we sometimes get the impression that he invented relativity out of nothing. On the contrary, one of Einstein’s motivations was to solve difficulties in knowing how different observers see magnetic and electric fields.

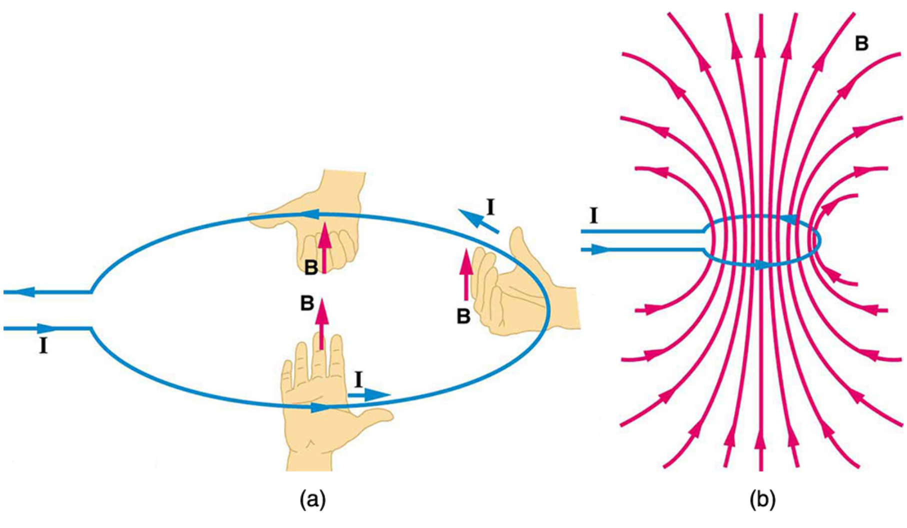
Magnetic Field Produced by a Current-Carrying Circular Loop

The magnetic field near a current-carrying loop of wire is shown in [link]. Both the direction and the magnitude of the magnetic field produced by a current-carrying loop are complex. RHR-2 can be used to give the direction of the field near the loop, but mapping with compasses and the rules about field lines given in Magnetic Fields and Magnetic Field Lines are needed for more detail. There is a simple formula for the **magnetic field strength at the center of a circular loop**. It is

Equation:

$$B = \frac{\mu_0 I}{2R} \text{ (at center of loop)},$$

where R is the radius of the loop. This equation is very similar to that for a straight wire, but it is valid *only* at the center of a circular loop of wire. The similarity of the equations does indicate that similar field strength can be obtained at the center of a loop. One way to get a larger field is to have N loops; then, the field is $B = N\mu_0 I/(2R)$. Note that the larger the loop, the smaller the field at its center, because the current is farther away.

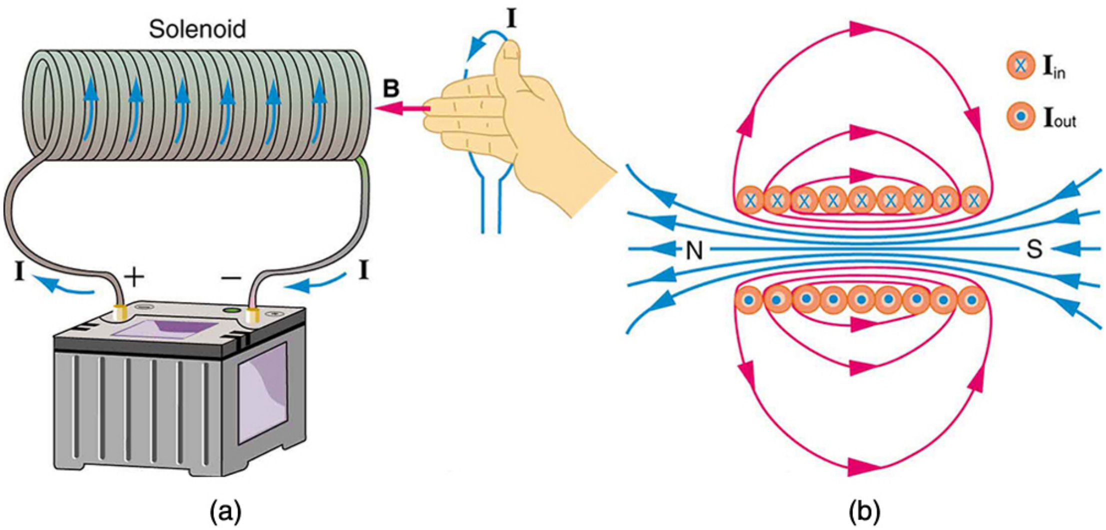


(a) RHR-2 gives the direction of the magnetic field inside and outside a current-carrying loop. (b) More detailed mapping with compasses or with a

Hall probe completes the picture. The field is similar to that of a bar magnet.

Magnetic Field Produced by a Current-Carrying Solenoid

A **solenoid** is a long coil of wire (with many turns or loops, as opposed to a flat loop). Because of its shape, the field inside a solenoid can be very uniform, and also very strong. The field just outside the coils is nearly zero. [link] shows how the field looks and how its direction is given by RHR-2.



(a) Because of its shape, the field inside a solenoid of length l is remarkably uniform in magnitude and direction, as indicated by the straight and uniformly spaced field lines. The field outside the coils is nearly zero. (b) This cutaway shows the magnetic field generated by the current in the solenoid.

The magnetic field inside of a current-carrying solenoid is very uniform in direction and magnitude. Only near the ends does it begin to weaken and change direction. The field outside has similar complexities to flat loops

and bar magnets, but the **magnetic field strength inside a solenoid** is simply

Equation:

$$B = \mu_0 \text{nI} \ \ (\text{inside a solenoid}),$$

where n is the number of loops per unit length of the solenoid $(n = N/l$, with N being the number of loops and l the length). Note that B is the field strength anywhere in the uniform region of the interior and not just at the center. Large uniform fields spread over a large volume are possible with solenoids, as [link] implies.

Example:

Calculating Field Strength inside a Solenoid

What is the field inside a 2.00-m-long solenoid that has 2000 loops and carries a 1600-A current?

Strategy

To find the field strength inside a solenoid, we use $B = \mu_0 \text{nI}$. First, we note the number of loops per unit length is

Equation:

$$n = \frac{N}{l} = \frac{2000}{2.00 \text{ m}} = 1000 \text{ m}^{-1} = 10 \text{ cm}^{-1}.$$

Solution

Substituting known values gives

Equation:

$$\begin{aligned} B &= \mu_0 \text{nI} = \left(4\pi \times 10^{-7} \text{ T} \cdot \text{m/A}\right)\left(1000 \text{ m}^{-1}\right)(1600 \text{ A}) \\ &= 2.01 \text{ T}. \end{aligned}$$

Discussion

This is a large field strength that could be established over a large-diameter solenoid, such as in medical uses of magnetic resonance imaging (MRI). The very large current is an indication that the fields of this strength are not

easily achieved, however. Such a large current through 1000 loops squeezed into a meter’s length would produce significant heating. Higher currents can be achieved by using superconducting wires, although this is expensive. There is an upper limit to the current, since the superconducting state is disrupted by very large magnetic fields.

There are interesting variations of the flat coil and solenoid. For example, the toroidal coil used to confine the reactive particles in tokamaks is much like a solenoid bent into a circle. The field inside a toroid is very strong but circular. Charged particles travel in circles, following the field lines, and collide with one another, perhaps inducing fusion. But the charged particles do not cross field lines and escape the toroid. A whole range of coil shapes are used to produce all sorts of magnetic field shapes. Adding ferromagnetic materials produces greater field strengths and can have a significant effect on the shape of the field. Ferromagnetic materials tend to trap magnetic fields (the field lines bend into the ferromagnetic material, leaving weaker fields outside it) and are used as shields for devices that are adversely affected by magnetic fields, including the Earth’s magnetic field.

Note:
PhET Explorations: Generator
Generate electricity with a bar magnet! Discover the physics behind the phenomena by exploring magnets and how you can use them to make a bulb light.

Generator

Section Summary

- The strength of the magnetic field created by current in a long straight wire is given by
  Equation:

  $$B = \frac{\mu_0 I}{2\pi r}(\text{long straight wire}),$$

  where I is the current, r is the shortest distance to the wire, and the constant $\mu_0 = 4\pi \times 10^{-7}\ \text{T} \cdot \text{m/A}$ is the permeability of free space.
- The direction of the magnetic field created by a long straight wire is given by right hand rule 2 (RHR-2): *Point the thumb of the right hand in the direction of current, and the fingers curl in the direction of the magnetic field loops* created by it.
- The magnetic field created by current following any path is the sum (or integral) of the fields due to segments along the path (magnitude and direction as for a straight wire), resulting in a general relationship between current and field known as Ampere’s law.
- The magnetic field strength at the center of a circular loop is given by
  Equation:

  $$B = \frac{\mu_0 I}{2R}(\text{at center of loop}),$$

  where R is the radius of the loop. This equation becomes $B = \mu_0 \text{nI}/(2R)$ for a flat coil of N loops. RHR-2 gives the direction of the field about the loop. A long coil is called a solenoid.
- The magnetic field strength inside a solenoid is
  Equation:

  $$B = \mu_0 \text{nI} \ \ (\text{inside a solenoid}),$$

  where n is the number of loops per unit length of the solenoid. The field inside is very uniform in magnitude and direction.

Conceptual Questions

Exercise:

Problem:

Make a drawing and use RHR-2 to find the direction of the magnetic field of a current loop in a motor (such as in [link]). Then show that the direction of the torque on the loop is the same as produced by like poles repelling and unlike poles attracting.

Glossary

right hand rule 2 (RHR-2)
: a rule to determine the direction of the magnetic field induced by a current-carrying wire: Point the thumb of the right hand in the direction of current, and the fingers curl in the direction of the magnetic field loops

magnetic field strength (magnitude) produced by a long straight current-carrying wire
: defined as $B = \frac{\mu_0 I}{2\pi r}$, where I is the current, r is the shortest distance to the wire, and μ_0 is the permeability of free space

permeability of free space
: the measure of the ability of a material, in this case free space, to support a magnetic field; the constant $\mu_0 = 4\pi \times 10^{-7}\ \mathrm{T \cdot m/A}$

magnetic field strength at the center of a circular loop
: defined as $B = \frac{\mu_0 I}{2R}$ where R is the radius of the loop

solenoid
: a thin wire wound into a coil that produces a magnetic field when an electric current is passed through it

magnetic field strength inside a solenoid
: defined as $B = \mu_0 n I$ where n is the number of loops per unit length of the solenoid ($n = N/l$, with N being the number of loops and l the length)

Biot-Savart law

a physical law that describes the magnetic field generated by an electric current in terms of a specific equation

Ampere's law

the physical law that states that the magnetic field around an electric current is proportional to the current; each segment of current produces a magnetic field like that of a long straight wire, and the total field of any shape current is the vector sum of the fields due to each segment

Maxwell's equations

a set of four equations that describe electromagnetic phenomena

Magnetic Force between Two Parallel Conductors

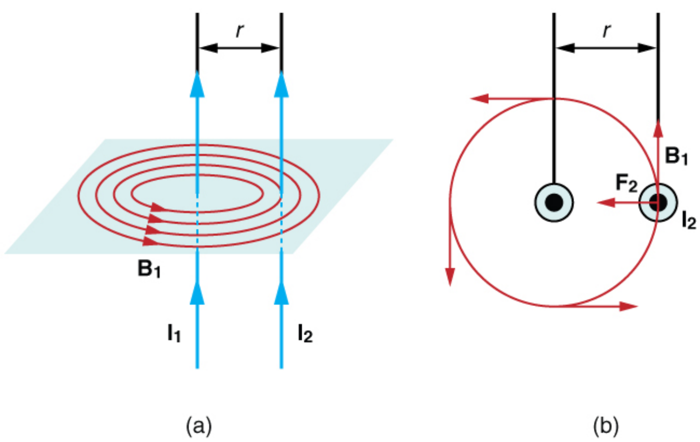
- Describe the effects of the magnetic force between two conductors.
- Calculate the force between two parallel conductors.

You might expect that there are significant forces between current-carrying wires, since ordinary currents produce significant magnetic fields and these fields exert significant forces on ordinary currents. But you might not expect that the force between wires is used to *define* the ampere. It might also surprise you to learn that this force has something to do with why large circuit breakers burn up when they attempt to interrupt large currents.

The force between two long straight and parallel conductors separated by a distance r can be found by applying what we have developed in preceding sections. [link] shows the wires, their currents, the fields they create, and the subsequent forces they exert on one another. Let us consider the field produced by wire 1 and the force it exerts on wire 2 (call the force F_2). The field due to I_1 at a distance r is given to be

Equation:

$$B_1 = \frac{\mu_0 I_1}{2\pi r}.$$



(a) The magnetic field produced by a long straight conductor is perpendicular to a parallel conductor, as indicated by RHR-2. (b) A view

from above of the two wires shown in (a), with one magnetic field line shown for each wire. RHR-1 shows that the force between the parallel conductors is attractive when the currents are in the same direction. A similar analysis shows that the force is repulsive between currents in opposite directions.

This field is uniform along wire 2 and perpendicular to it, and so the force F_2 it exerts on wire 2 is given by $F = \text{IlB}\ \sin\theta$ with $\sin\theta = 1$:

Equation:

$$F_2 = I_2 l B_1.$$

By Newton's third law, the forces on the wires are equal in magnitude, and so we just write F for the magnitude of F_2. (Note that $F_1 = -F_2$.) Since the wires are very long, it is convenient to think in terms of F/l, the force per unit length. Substituting the expression for B_1 into the last equation and rearranging terms gives

Equation:

$$\frac{F}{l} = \frac{\mu_0 I_1 I_2}{2\pi r}.$$

F/l is the force per unit length between two parallel currents I_1 and I_2 separated by a distance r. The force is attractive if the currents are in the same direction and repulsive if they are in opposite directions.

This force is responsible for the *pinch effect* in electric arcs and plasmas. The force exists whether the currents are in wires or not. In an electric arc, where currents are moving parallel to one another, there is an attraction that squeezes currents into a smaller tube. In large circuit breakers, like those

used in neighborhood power distribution systems, the pinch effect can concentrate an arc between plates of a switch trying to break a large current, burn holes, and even ignite the equipment. Another example of the pinch effect is found in the solar plasma, where jets of ionized material, such as solar flares, are shaped by magnetic forces.

The *operational definition of the ampere* is based on the force between current-carrying wires. Note that for parallel wires separated by 1 meter with each carrying 1 ampere, the force per meter is

Equation:

$$\frac{F}{l} = \frac{\left(4\pi \times 10^{-7}\ \mathrm{T \cdot m/A}\right)(1\ \mathrm{A})^2}{(2\pi)(1\ \mathrm{m})} = 2 \times 10^{-7}\ \mathrm{N/m}.$$

Since μ_0 is exactly $4\pi \times 10^{-7}\ \mathrm{T \cdot m/A}$ by definition, and because $1\ \mathrm{T} = 1\ \mathrm{N/(A \cdot m)}$, the force per meter is exactly $2 \times 10^{-7}\ \mathrm{N/m}$. This is the basis of the operational definition of the ampere.

Note:

The Ampere

The official definition of the ampere is:

One ampere of current through each of two parallel conductors of infinite length, separated by one meter in empty space free of other magnetic fields, causes a force of exactly $2 \times 10^{-7}\ \mathrm{N/m}$ on each conductor.

Infinite-length straight wires are impractical and so, in practice, a current balance is constructed with coils of wire separated by a few centimeters. Force is measured to determine current. This also provides us with a method for measuring the coulomb. We measure the charge that flows for a current of one ampere in one second. That is, $1\ \mathrm{C} = 1\ \mathrm{A \cdot s}$. For both the ampere and the coulomb, the method of measuring force between conductors is the most accurate in practice.

Section Summary

- The force between two parallel currents I_1 and I_2, separated by a distance r, has a magnitude per unit length given by
  Equation:
  $$\frac{F}{l} = \frac{\mu_0 I_1 I_2}{2\pi r}.$$
- The force is attractive if the currents are in the same direction, repulsive if they are in opposite directions.

Conceptual Questions

Exercise:

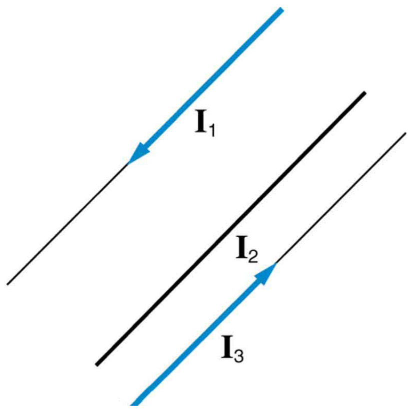
Problem:

Is the force attractive or repulsive between the hot and neutral lines hung from power poles? Why?

Exercise:

Problem:

If you have three parallel wires in the same plane, as in [link], with currents in the outer two running in opposite directions, is it possible for the middle wire to be repelled by both? Attracted by both? Explain.



Three parallel

coplanar wires with currents in the outer two in opposite directions.

Exercise:

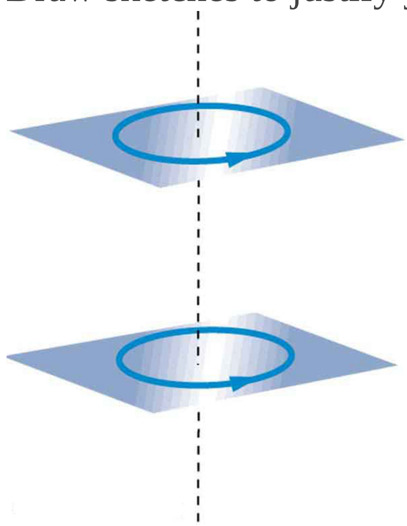
Problem:

Suppose two long straight wires run perpendicular to one another without touching. Does one exert a net force on the other? If so, what is its direction? Does one exert a net torque on the other? If so, what is its direction? Justify your responses by using the right hand rules.

Exercise:

Problem:

Use the right hand rules to show that the force between the two loops in [link] is attractive if the currents are in the same direction and repulsive if they are in opposite directions. Is this consistent with like poles of the loops repelling and unlike poles of the loops attracting? Draw sketches to justify your answers.



Two loops of wire carrying currents

can exert forces
and torques on one
another.

Exercise:

Problem:

If one of the loops in [link] is tilted slightly relative to the other and their currents are in the same direction, what are the directions of the torques they exert on each other? Does this imply that the poles of the bar magnet-like fields they create will line up with each other if the loops are allowed to rotate?

Exercise:

Problem:

Electric field lines can be shielded by the Faraday cage effect. Can we have magnetic shielding? Can we have gravitational shielding?

Problems & Exercises

Exercise:

Problem:

(a) The hot and neutral wires supplying DC power to a light-rail commuter train carry 800 A and are separated by 75.0 cm. What is the magnitude and direction of the force between 50.0 m of these wires? (b) Discuss the practical consequences of this force, if any.

Solution:

(a) 8.53 N, repulsive

(b) This force is repulsive and therefore there is never a risk that the two wires will touch and short circuit.

Exercise:

Problem:

The force per meter between the two wires of a jumper cable being used to start a stalled car is 0.225 N/m. (a) What is the current in the wires, given they are separated by 2.00 cm? (b) Is the force attractive or repulsive?

Exercise:

Problem:

A 2.50-m segment of wire supplying current to the motor of a submerged submarine carries 1000 A and feels a 4.00-N repulsive force from a parallel wire 5.00 cm away. What is the direction and magnitude of the current in the other wire?

Solution:

400 A in the opposite direction

Exercise:

Problem:

The wire carrying 400 A to the motor of a commuter train feels an attractive force of $4.00 \times 10^{-3}\ \mathrm{N/m}$ due to a parallel wire carrying 5.00 A to a headlight. (a) How far apart are the wires? (b) Are the currents in the same direction?

Exercise:

Problem:

An AC appliance cord has its hot and neutral wires separated by 3.00 mm and carries a 5.00-A current. (a) What is the average force per meter between the wires in the cord? (b) What is the maximum force per meter between the wires? (c) Are the forces attractive or repulsive? (d) Do appliance cords need any special design features to compensate for these forces?

Solution:

(a) $1.67 \times 10^{-3}\ \mathrm{N/m}$

(b) $3.33 \times 10^{-3}\ \mathrm{N/m}$

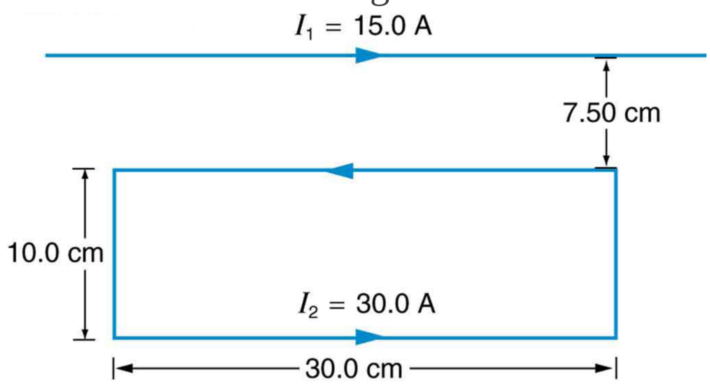
(c) Repulsive

(d) No, these are very small forces

Exercise:

Problem:

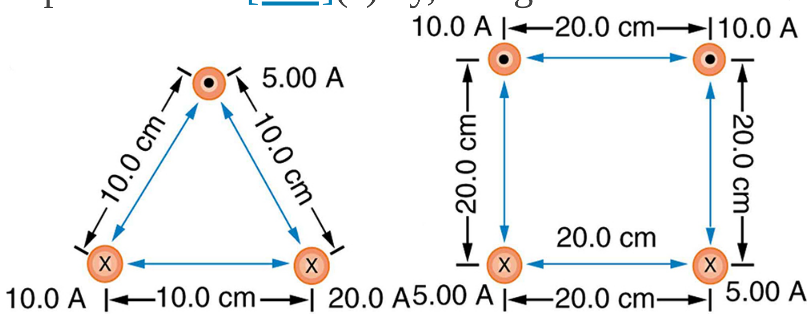
[link] shows a long straight wire near a rectangular current loop. What is the direction and magnitude of the total force on the loop?



Exercise:

Problem:

Find the direction and magnitude of the force that each wire experiences in [link](a) by, using vector addition.



Solution:

(a) Top wire: $2.65 \times 10^{-4}\ \mathrm{N/m}$ s, 10.9^{o} to left of up

(b) Lower left wire: $3.61 \times 10^{-4}\ \mathrm{N/m}$, 13.9^{o} down from right

(c) Lower right wire: $3.46 \times 10^{-4}\ \mathrm{N/m}$, 30.0^{o} down from left

Exercise:

Problem:

Find the direction and magnitude of the force that each wire experiences in [link](b), using vector addition.

More Applications of Magnetism

- Describe some applications of magnetism.

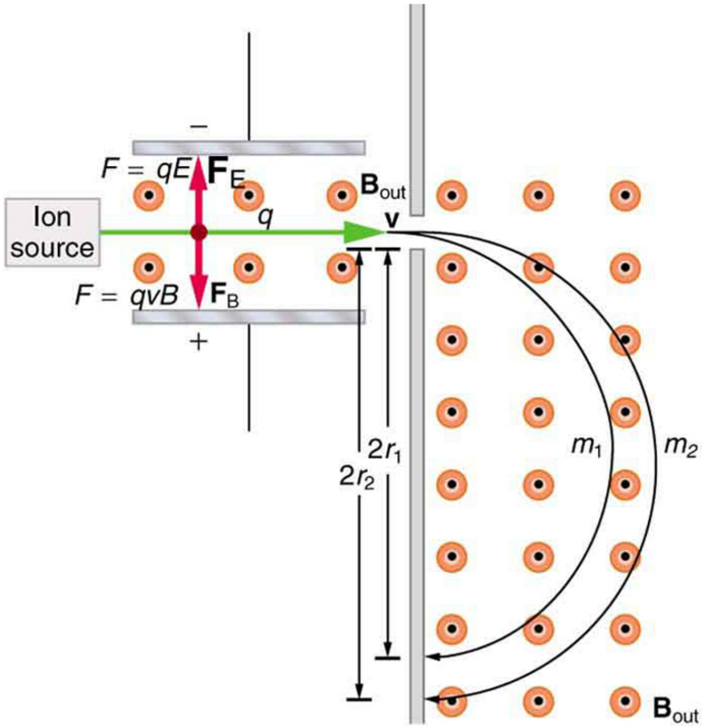
Mass Spectrometry

The curved paths followed by charged particles in magnetic fields can be put to use. A charged particle moving perpendicular to a magnetic field travels in a circular path having a radius r.

Equation:

$$r = \frac{\mathrm{mv}}{\mathrm{qB}}$$

It was noted that this relationship could be used to measure the mass of charged particles such as ions. A mass spectrometer is a device that measures such masses. Most mass spectrometers use magnetic fields for this purpose, although some of them have extremely sophisticated designs. Since there are five variables in the relationship, there are many possibilities. However, if v, q, and B can be fixed, then the radius of the path r is simply proportional to the mass m of the charged particle. Let us examine one such mass spectrometer that has a relatively simple design. (See [link].) The process begins with an ion source, a device like an electron gun. The ion source gives ions their charge, accelerates them to some velocity v, and directs a beam of them into the next stage of the spectrometer. This next region is a *velocity selector* that only allows particles with a particular value of v to get through.



This mass spectrometer uses a velocity selector to fix v so that the radius of the path is proportional to mass.

The velocity selector has both an electric field and a magnetic field, perpendicular to one another, producing forces in opposite directions on the ions. Only those ions for which the forces balance travel in a straight line into the next region. If the forces balance, then the electric force $F = \text{qE}$ equals the magnetic force $F = \text{qvB}$, so that $\text{qE} = \text{qvB}$. Noting that q cancels, we see that

Equation:

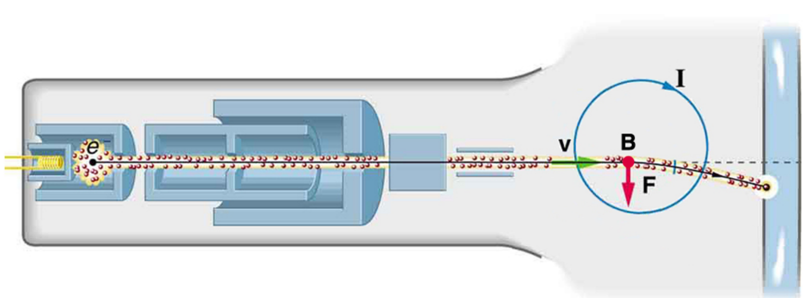
$$v = \frac{E}{B}$$

is the velocity particles must have to make it through the velocity selector, and further, that v can be selected by varying E and B. In the final region, there is only a uniform magnetic field, and so the charged particles move in circular arcs with radii proportional to particle mass. The paths also depend on charge q, but since q is in multiples of electron charges, it is easy to determine and to discriminate between ions in different charge states.

Mass spectrometry today is used extensively in chemistry and biology laboratories to identify chemical and biological substances according to their mass-to-charge ratios. In medicine, mass spectrometers are used to measure the concentration of isotopes used as tracers. Usually, biological molecules such as proteins are very large, so they are broken down into smaller fragments before analyzing. Recently, large virus particles have been analyzed as a whole on mass spectrometers. Sometimes a gas chromatograph or high-performance liquid chromatograph provides an initial separation of the large molecules, which are then input into the mass spectrometer.

Cathode Ray Tubes—CRTs—and the Like

What do non-flat-screen TVs, old computer monitors, x-ray machines, and the 2-mile-long Stanford Linear Accelerator have in common? All of them accelerate electrons, making them different versions of the electron gun. Many of these devices use magnetic fields to steer the accelerated electrons. [link] shows the construction of the type of cathode ray tube (CRT) found in some TVs, oscilloscopes, and old computer monitors. Two pairs of coils are used to steer the electrons, one vertically and the other horizontally, to their desired destination.



The cathode ray tube (CRT) is so named because rays of electrons originate at the cathode in the electron gun. Magnetic coils are used to steer the beam in many CRTs. In this case, the beam is moved down. Another pair of horizontal coils would steer the beam horizontally.

Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is one of the most useful and rapidly growing medical imaging tools. It non-invasively produces two-dimensional and three-dimensional images of the body that provide important medical information with none of the hazards of x-rays. MRI is based on an effect called **nuclear magnetic resonance (NMR)** in which an externally applied magnetic field interacts with the nuclei of certain atoms, particularly those of hydrogen (protons). These nuclei possess their own small magnetic fields, similar to those of electrons and the current loops discussed earlier in this chapter.

When placed in an external magnetic field, such nuclei experience a torque that pushes or aligns the nuclei into one of two new energy states—depending on the orientation of its spin (analogous to the N pole and S pole in a bar magnet). Transitions from the lower to higher energy state can be achieved by using an external radio frequency signal to “flip” the orientation of the small magnets. (This is actually a quantum mechanical process. The direction of the nuclear magnetic field is quantized as is energy in the radio waves. We will return to these topics in later chapters.) The specific frequency of the radio waves that are absorbed and reemitted depends sensitively on the type of nucleus, the chemical environment, and the external magnetic field strength. Therefore, this is a *resonance* phenomenon in which *nuclei* in a *magnetic* field act like resonators (analogous to those discussed in the treatment of sound in Oscillatory

Motion and Waves) that absorb and reemit only certain frequencies. Hence, the phenomenon is named *nuclear magnetic resonance (NMR).*

NMR has been used for more than 50 years as an analytical tool. It was formulated in 1946 by F. Bloch and E. Purcell, with the 1952 Nobel Prize in Physics going to them for their work. Over the past two decades, NMR has been developed to produce detailed images in a process now called magnetic resonance imaging (MRI), a name coined to avoid the use of the word “nuclear” and the concomitant implication that nuclear radiation is involved. (It is not.) The 2003 Nobel Prize in Medicine went to P. Lauterbur and P. Mansfield for their work with MRI applications.

The largest part of the MRI unit is a superconducting magnet that creates a magnetic field, typically between 1 and 2 T in strength, over a relatively large volume. MRI images can be both highly detailed and informative about structures and organ functions. It is helpful that normal and non-normal tissues respond differently for slight changes in the magnetic field. In most medical images, the protons that are hydrogen nuclei are imaged. (About 2/3 of the atoms in the body are hydrogen.) Their location and density give a variety of medically useful information, such as organ function, the condition of tissue (as in the brain), and the shape of structures, such as vertebral disks and knee-joint surfaces. MRI can also be used to follow the movement of certain ions across membranes, yielding information on active transport, osmosis, dialysis, and other phenomena. With excellent spatial resolution, MRI can provide information about tumors, strokes, shoulder injuries, infections, etc.

An image requires position information as well as the density of a nuclear type (usually protons). By varying the magnetic field slightly over the volume to be imaged, the resonant frequency of the protons is made to vary with position. Broadcast radio frequencies are swept over an appropriate range and nuclei absorb and reemit them only if the nuclei are in a magnetic field with the correct strength. The imaging receiver gathers information through the body almost point by point, building up a tissue map. The reception of reemitted radio waves as a function of frequency thus gives position information. These “slices” or cross sections through the body are only several mm thick. The intensity of the reemitted radio waves is

proportional to the concentration of the nuclear type being flipped, as well as information on the chemical environment in that area of the body. Various techniques are available for enhancing contrast in images and for obtaining more information. Scans called T1, T2, or proton density scans rely on different relaxation mechanisms of nuclei. Relaxation refers to the time it takes for the protons to return to equilibrium after the external field is turned off. This time depends upon tissue type and status (such as inflammation).

While MRI images are superior to x rays for certain types of tissue and have none of the hazards of x rays, they do not completely supplant x-ray images. MRI is less effective than x rays for detecting breaks in bone, for example, and in imaging breast tissue, so the two diagnostic tools complement each other. MRI images are also expensive compared to simple x-ray images and tend to be used most often where they supply information not readily obtained from x rays. Another disadvantage of MRI is that the patient is totally enclosed with detectors close to the body for about 30 minutes or more, leading to claustrophobia. It is also difficult for the obese patient to be in the magnet tunnel. New “open-MRI” machines are now available in which the magnet does not completely surround the patient.

Over the last decade, the development of much faster scans, called “functional MRI” (fMRI), has allowed us to map the functioning of various regions in the brain responsible for thought and motor control. This technique measures the change in blood flow for activities (thought, experiences, action) in the brain. The nerve cells increase their consumption of oxygen when active. Blood hemoglobin releases oxygen to active nerve cells and has somewhat different magnetic properties when oxygenated than when deoxygenated. With MRI, we can measure this and detect a blood oxygen-dependent signal. Most of the brain scans today use fMRI.

Other Medical Uses of Magnetic Fields

Currents in nerve cells and the heart create magnetic fields like any other currents. These can be measured but with some difficulty since their strengths are about 10^{-6} to 10^{-8} *less* than the Earth’s magnetic field. Recording of the heart’s magnetic field as it beats is called a

magnetocardiogram (MCG), while measurements of the brain’s magnetic field is called a **magnetoencephalogram (MEG)**. Both give information that differs from that obtained by measuring the electric fields of these organs (ECGs and EEGs), but they are not yet of sufficient importance to make these difficult measurements common.

In both of these techniques, the sensors do not touch the body. MCG can be used in fetal studies, and is probably more sensitive than echocardiography. MCG also looks at the heart’s electrical activity whose voltage output is too small to be recorded by surface electrodes as in EKG. It has the potential of being a rapid scan for early diagnosis of cardiac ischemia (obstruction of blood flow to the heart) or problems with the fetus.

MEG can be used to identify abnormal electrical discharges in the brain that produce weak magnetic signals. Therefore, it looks at brain activity, not just brain structure. It has been used for studies of Alzheimer’s disease and epilepsy. Advances in instrumentation to measure very small magnetic fields have allowed these two techniques to be used more in recent years. What is used is a sensor called a SQUID, for superconducting quantum interference device. This operates at liquid helium temperatures and can measure magnetic fields thousands of times smaller than the Earth’s.

Finally, there is a burgeoning market for magnetic cures in which magnets are applied in a variety of ways to the body, from magnetic bracelets to magnetic mattresses. The best that can be said for such practices is that they are apparently harmless, unless the magnets get close to the patient’s computer or magnetic storage disks. Claims are made for a broad spectrum of benefits from cleansing the blood to giving the patient more energy, but clinical studies have not verified these claims, nor is there an identifiable mechanism by which such benefits might occur.

Note:
PhET Explorations: Magnet and Compass
Ever wonder how a compass worked to point you to the Arctic? Explore the interactions between a compass and bar magnet, and then add the Earth and find the surprising answer! Vary the magnet's strength, and see how

things change both inside and outside. Use the field meter to measure how the magnetic field changes.
https://archive.cnx.org/specials/5ca3e2cc-ae74-11e5-b6d3-f3c228f04b5c/magnet-and-compass/#sim-bar-magnet

Section Summary

- Crossed (perpendicular) electric and magnetic fields act as a velocity filter, giving equal and opposite forces on any charge with velocity perpendicular to the fields and of magnitude
  Equation:

$$v = \frac{E}{B}.$$

Conceptual Questions

Exercise:

Problem:

Measurements of the weak and fluctuating magnetic fields associated with brain activity are called magnetoencephalograms (MEGs). Do the brain’s magnetic fields imply coordinated or uncoordinated nerve impulses? Explain.

Exercise:

Problem:

Discuss the possibility that a Hall voltage would be generated on the moving heart of a patient during MRI imaging. Also discuss the same effect on the wires of a pacemaker. (The fact that patients with pacemakers are not given MRIs is significant.)

Exercise:

Problem:

A patient in an MRI unit turns his head quickly to one side and experiences momentary dizziness and a strange taste in his mouth. Discuss the possible causes.

Exercise:

Problem:

You are told that in a certain region there is either a uniform electric or magnetic field. What measurement or observation could you make to determine the type? (Ignore the Earth’s magnetic field.)

Exercise:

Problem:

An example of magnetohydrodynamics (MHD) comes from the flow of a river (salty water). This fluid interacts with the Earth’s magnetic field to produce a potential difference between the two river banks. How would you go about calculating the potential difference?

Exercise:

Problem:

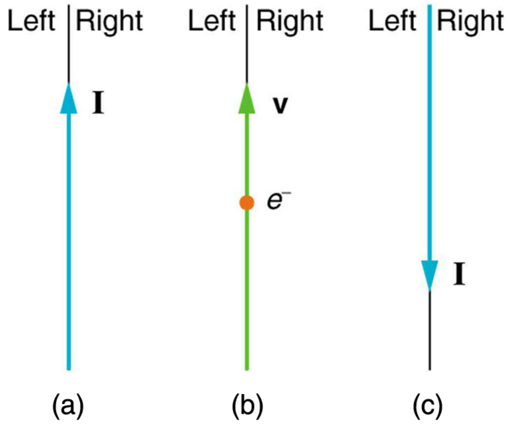
Draw gravitational field lines between 2 masses, electric field lines between a positive and a negative charge, electric field lines between 2 positive charges and magnetic field lines around a magnet. Qualitatively describe the differences between the fields and the entities responsible for the field lines.

Problems & Exercises

Exercise:

Problem:

Indicate whether the magnetic field created in each of the three situations shown in [link] is into or out of the page on the left and right of the current.



Solution:

(a) right-into page, left-out of page

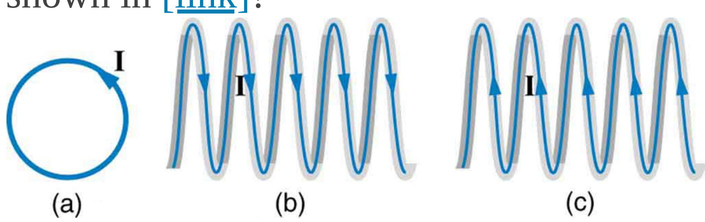
(b) right-out of page, left-into page

(c) right-out of page, left-into page

Exercise:

Problem:

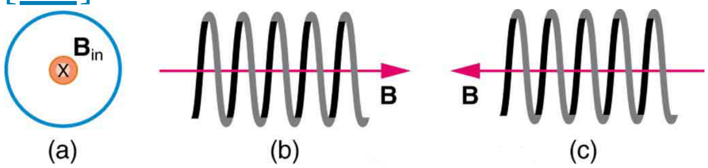
What are the directions of the fields in the center of the loop and coils shown in [link]?



Exercise:

Problem:

What are the directions of the currents in the loop and coils shown in [link]?



Solution:

(a) clockwise

(b) clockwise as seen from the left

(c) clockwise as seen from the right

Exercise:

Problem:

To see why an MRI utilizes iron to increase the magnetic field created by a coil, calculate the current needed in a 400-loop-per-meter circular coil 0.660 m in radius to create a 1.20-T field (typical of an MRI instrument) at its center with no iron present. The magnetic field of a proton is approximately like that of a circular current loop 0.650×10^{-15} m in radius carrying 1.05×10^{4} A. What is the field at the center of such a loop?

Solution:

1.01×10^{13} T

Exercise:

Problem:

Inside a motor, 30.0 A passes through a 250-turn circular loop that is 10.0 cm in radius. What is the magnetic field strength created at its center?

Exercise:

Problem:

Nonnuclear submarines use batteries for power when submerged. (a) Find the magnetic field 50.0 cm from a straight wire carrying 1200 A from the batteries to the drive mechanism of a submarine. (b) What is the field if the wires to and from the drive mechanism are side by side? (c) Discuss the effects this could have for a compass on the submarine that is not shielded.

Solution:

(a) $4.80 \times 10^{-4}\ \mathrm{T}$

(b) Zero

(c) If the wires are not paired, the field is about 10 times stronger than Earth's magnetic field and so could severely disrupt the use of a compass.

Exercise:

Problem:

How strong is the magnetic field inside a solenoid with 10,000 turns per meter that carries 20.0 A?

Exercise:

Problem:

What current is needed in the solenoid described in [link] to produce a magnetic field 10^4 times the Earth's magnetic field of $5.00 \times 10^{-5}\ \mathrm{T}$?

Solution:

39.8 A

Exercise:

Problem:

How far from the starter cable of a car, carrying 150 A, must you be to experience a field less than the Earth's $(5.00 \times 10^{-5}\ \mathrm{T})$? Assume a long straight wire carries the current. (In practice, the body of your car shields the dashboard compass.)

Exercise:

Problem:

Measurements affect the system being measured, such as the current loop in [link]. (a) Estimate the field the loop creates by calculating the field at the center of a circular loop 20.0 cm in diameter carrying 5.00 A. (b) What is the smallest field strength this loop can be used to measure, if its field must alter the measured field by less than 0.0100%?

Solution:

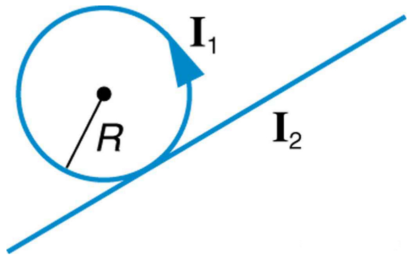
(a) $3.14 \times 10^{-5}\ \mathrm{T}$

(b) 0.314 T

Exercise:

Problem:

[link] shows a long straight wire just touching a loop carrying a current I_1. Both lie in the same plane. (a) What direction must the current I_2 in the straight wire have to create a field at the center of the loop in the direction opposite to that created by the loop? (b) What is the ratio of I_1/I_2 that gives zero field strength at the center of the loop? (c) What is the direction of the field directly above the loop under this circumstance?



Exercise:

Problem:

Find the magnitude and direction of the magnetic field at the point equidistant from the wires in [link](a), using the rules of vector addition to sum the contributions from each wire.

Solution:

7.55×10^{-5} T, 23.4º

Exercise:

Problem:

Find the magnitude and direction of the magnetic field at the point equidistant from the wires in [link](b), using the rules of vector addition to sum the contributions from each wire.

Exercise:

Problem:

What current is needed in the top wire in [link](a) to produce a field of zero at the point equidistant from the wires, if the currents in the bottom two wires are both 10.0 A into the page?

Solution:

10.0 A

Exercise:

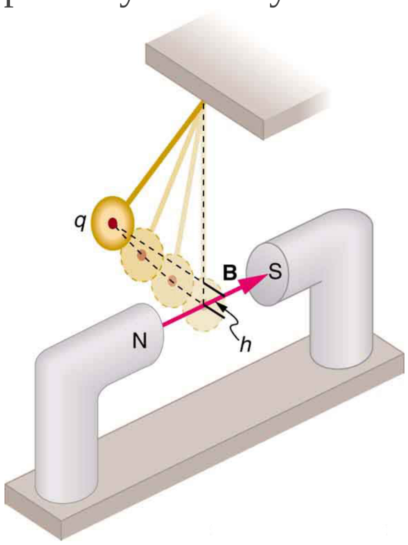
Problem:

Calculate the size of the magnetic field 20 m below a high voltage power line. The line carries 450 MW at a voltage of 300,000 V.

Exercise:

Problem: Integrated Concepts

(a) A pendulum is set up so that its bob (a thin copper disk) swings between the poles of a permanent magnet as shown in [link]. What is the magnitude and direction of the magnetic force on the bob at the lowest point in its path, if it has a positive $0.250\ \mu C$ charge and is released from a height of 30.0 cm above its lowest point? The magnetic field strength is 1.50 T. (b) What is the acceleration of the bob at the bottom of its swing if its mass is 30.0 grams and it is hung from a flexible string? Be certain to include a free-body diagram as part of your analysis.



Solution:

(a) 9.09×10^{-7} N upward

(b) $3.03 \times 10^{-5}\ \mathrm{m/s^2}$

Exercise:

Problem: Integrated Concepts

(a) What voltage will accelerate electrons to a speed of $6.00 \times 10^{-7}\ \mathrm{m/s}$? (b) Find the radius of curvature of the path of a *proton* accelerated through this potential in a 0.500-T field and compare this with the radius of curvature of an electron accelerated through the same potential.

Exercise:

Problem: Integrated Concepts

Find the radius of curvature of the path of a 25.0-MeV proton moving perpendicularly to the 1.20-T field of a cyclotron.

Solution:

60.2 cm

Exercise:

Problem: Integrated Concepts

To construct a nonmechanical water meter, a 0.500-T magnetic field is placed across the supply water pipe to a home and the Hall voltage is recorded. (a) Find the flow rate in liters per second through a 3.00-cm-diameter pipe if the Hall voltage is 60.0 mV. (b) What would the Hall voltage be for the same flow rate through a 10.0-cm-diameter pipe with the same field applied?

Exercise:

Problem: Integrated Concepts

(a) Using the values given for an MHD drive in [link], and assuming the force is uniformly applied to the fluid, calculate the pressure created in $\mathrm{N/m^2}$. (b) Is this a significant fraction of an atmosphere?

Solution:

(a) $1.02 \times 10^3\ \mathrm{N/m}^2$

(b) Not a significant fraction of an atmosphere

Exercise:

Problem: Integrated Concepts

(a) Calculate the maximum torque on a 50-turn, 1.50 cm radius circular current loop carrying $50\ \mu\mathrm{A}$ in a 0.500-T field. (b) If this coil is to be used in a galvanometer that reads $50\ \mu\mathrm{A}$ full scale, what force constant spring must be used, if it is attached 1.00 cm from the axis of rotation and is stretched by the 60^{o} arc moved?

Exercise:

Problem: Integrated Concepts

A current balance used to define the ampere is designed so that the current through it is constant, as is the distance between wires. Even so, if the wires change length with temperature, the force between them will change. What percent change in force per degree will occur if the wires are copper?

Solution:

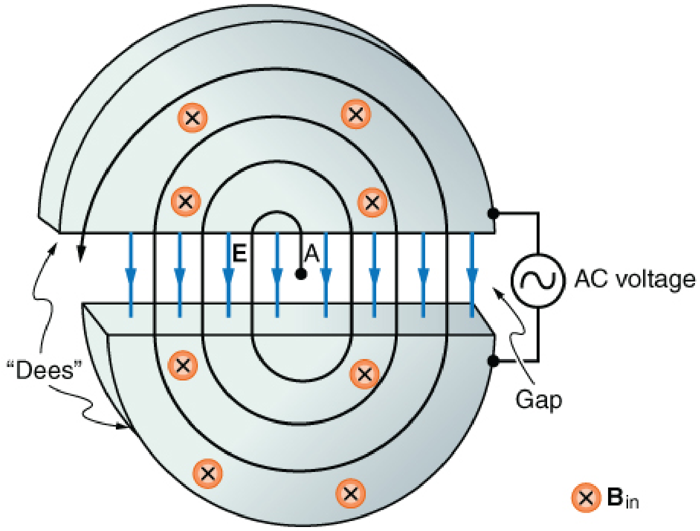
$17.0 \times 10^{-4}\%/^{\mathrm{o}}\mathrm{C}$

Exercise:

Problem:Integrated Concepts

(a) Show that the period of the circular orbit of a charged particle moving perpendicularly to a uniform magnetic field is $T = 2\pi m/(\mathrm{qB})$. (b) What is the frequency f? (c) What is the angular

velocity ω? Note that these results are independent of the velocity and radius of the orbit and, hence, of the energy of the particle. ([link].)



Cyclotrons accelerate charged particles orbiting in a magnetic field by placing an AC voltage on the metal Dees, between which the particles move, so that energy is added twice each orbit. The frequency is constant, since it is independent of the particle energy—the radius of the orbit simply increases with energy until the particles approach the edge and are extracted for various experiments and applications.

Exercise:

Problem: Integrated Concepts

A cyclotron accelerates charged particles as shown in [link]. Using the results of the previous problem, calculate the frequency of the accelerating voltage needed for a proton in a 1.20-T field.

Solution:

18.3 MHz

Exercise:

Problem: Integrated Concepts

(a) A 0.140-kg baseball, pitched at 40.0 m/s horizontally and perpendicular to the Earth's horizontal $5.00 \times 10^{-5}\ \mathrm{T}$ field, has a 100-nC charge on it. What distance is it deflected from its path by the magnetic force, after traveling 30.0 m horizontally? (b) Would you suggest this as a secret technique for a pitcher to throw curve balls?

Exercise:

Problem: Integrated Concepts

(a) What is the direction of the force on a wire carrying a current due east in a location where the Earth's field is due north? Both are parallel to the ground. (b) Calculate the force per meter if the wire carries 20.0 A and the field strength is $3.00 \times 10^{-5}\ \mathrm{T}$. (c) What diameter copper wire would have its weight supported by this force? (d) Calculate the resistance per meter and the voltage per meter needed.

Solution:

(a) Straight up

(b) $6.00 \times 10^{-4}\ \mathrm{N/m}$

(c) 94.1 μm

(d)2.47 Ω/m, 49.4 V/m

Exercise:

Problem: Integrated Concepts

One long straight wire is to be held directly above another by repulsion between their currents. The lower wire carries 100 A and the wire 7.50 cm above it is 10-gauge (2.588 mm diameter) copper wire. (a) What current must flow in the upper wire, neglecting the Earth's field? (b) What is the smallest current if the Earth's $3.00 \times 10^{-5}\ \mathrm{T}$ field is parallel to the ground and is not neglected? (c) Is the supported wire in a stable or unstable equilibrium if displaced vertically? If displaced horizontally?

Exercise:

Problem: Unreasonable Results

(a) Find the charge on a baseball, thrown at 35.0 m/s perpendicular to the Earth's $5.00 \times 10^{-5}\ \mathrm{T}$ field, that experiences a 1.00-N magnetic force. (b) What is unreasonable about this result? (c) Which assumption or premise is responsible?

Solution:

(a) 571 C

(b) Impossible to have such a large separated charge on such a small object.

(c) The 1.00-N force is much too great to be realistic in the Earth's field.

Exercise:

Problem: Unreasonable Results

A charged particle having mass 6.64×10^{-27} kg (that of a helium atom) moving at $8.70 \times 10^{5}\ \mathrm{m/s}$ perpendicular to a 1.50-T magnetic field travels in a circular path of radius 16.0 mm. (a) What is the charge of the particle? (b) What is unreasonable about this result? (c) Which assumptions are responsible?

Exercise:

Problem:Unreasonable Results

An inventor wants to generate 120-V power by moving a 1.00-m-long wire perpendicular to the Earth’s 5.00×10^{-5} T field. (a) Find the speed with which the wire must move. (b) What is unreasonable about this result? (c) Which assumption is responsible?

Solution:

(a) 2.40×10^{6} m/s

(b) The speed is too high to be practical $\leq$ 1% speed of light

(c) The assumption that you could reasonably generate such a voltage with a single wire in the Earth’s field is unreasonable

Exercise:

Problem: Unreasonable Results

Frustrated by the small Hall voltage obtained in blood flow measurements, a medical physicist decides to increase the applied magnetic field strength to get a 0.500-V output for blood moving at 30.0 cm/s in a 1.50-cm-diameter vessel. (a) What magnetic field strength is needed? (b) What is unreasonable about this result? (c) Which premise is responsible?

Exercise:

Problem:Unreasonable Results

A surveyor 100 m from a long straight 200-kV DC power line suspects that its magnetic field may equal that of the Earth and affect compass readings. (a) Calculate the current in the wire needed to create a 5.00×10^{-5} T field at this distance. (b) What is unreasonable about this result? (c) Which assumption or premise is responsible?

Solution:

(a) 25.0 kA

(b) This current is unreasonably high. It implies a total power delivery in the line of 50.0x10^9 W, which is much too high for standard transmission lines.

(c)100 meters is a long distance to obtain the required field strength. Also coaxial cables are used for transmission lines so that there is virtually no field for DC power lines, because of cancellation from opposing currents. The surveyor's concerns are not a problem for his magnetic field measurements.

Exercise:

Problem: **Construct Your Own Problem**

Consider a mass separator that applies a magnetic field perpendicular to the velocity of ions and separates the ions based on the radius of curvature of their paths in the field. Construct a problem in which you calculate the magnetic field strength needed to separate two ions that differ in mass, but not charge, and have the same initial velocity. Among the things to consider are the types of ions, the velocities they can be given before entering the magnetic field, and a reasonable value for the radius of curvature of the paths they follow. In addition, calculate the separation distance between the ions at the point where they are detected.

Exercise:

Problem: **Construct Your Own Problem**

Consider using the torque on a current-carrying coil in a magnetic field to detect relatively small magnetic fields (less than the field of the Earth, for example). Construct a problem in which you calculate the maximum torque on a current-carrying loop in a magnetic field. Among the things to be considered are the size of the coil, the number

of loops it has, the current you pass through the coil, and the size of the field you wish to detect. Discuss whether the torque produced is large enough to be effectively measured. Your instructor may also wish for you to consider the effects, if any, of the field produced by the coil on the surroundings that could affect detection of the small field.

Glossary

magnetic resonance imaging (MRI)
: a medical imaging technique that uses magnetic fields create detailed images of internal tissues and organs

nuclear magnetic resonance (NMR)
: a phenomenon in which an externally applied magnetic field interacts with the nuclei of certain atoms

magnetocardiogram (MCG)
: a recording of the heart's magnetic field as it beats

magnetoencephalogram (MEG)
: a measurement of the brain's magnetic field

Introduction to Electromagnetic Waves
class="introduction"

Human eyes detect these orange “sea goldie” fish swimming over a coral reef in the blue waters of the Gulf of Eilat (Red Sea) using visible light. (credit: Daviddarom, Wikimedia Commons)

The beauty of a coral reef, the warm radiance of sunshine, the sting of sunburn, the X-ray revealing a broken bone, even microwave popcorn—all are brought to us by **electromagnetic waves**. The list of the various types of electromagnetic waves, ranging from radio transmission waves to nuclear gamma-ray (γ-ray) emissions, is interesting in itself.

Even more intriguing is that all of these widely varied phenomena are different manifestations of the same thing—electromagnetic waves. (See [link].) What are electromagnetic waves? How are they created, and how do they travel? How can we understand and organize their widely varying properties? What is their relationship to electric and magnetic effects? These and other questions will be explored.

Note:
Misconception Alert: Sound Waves vs. Radio Waves
Many people confuse sound waves with **radio waves**, one type of electromagnetic (EM) wave. However, sound and radio waves are

completely different phenomena. Sound creates pressure variations (waves) in matter, such as air or water, or your eardrum. Conversely, radio waves are *electromagnetic waves*, like visible light, infrared, ultraviolet, X-rays, and gamma rays. EM waves don’t need a medium in which to propagate; they can travel through a vacuum, such as outer space.
A radio works because sound waves played by the D.J. at the radio station are converted into electromagnetic waves, then encoded and transmitted in the radio-frequency range. The radio in your car receives the radio waves, decodes the information, and uses a speaker to change it back into a sound wave, bringing sweet music to your ears.

Discovering a New Phenomenon

It is worth noting at the outset that the general phenomenon of electromagnetic waves was predicted by theory before it was realized that light is a form of electromagnetic wave. The prediction was made by James Clerk Maxwell in the mid-19th century when he formulated a single theory combining all the electric and magnetic effects known by scientists at that time. “Electromagnetic waves” was the name he gave to the phenomena his theory predicted.

Such a theoretical prediction followed by experimental verification is an indication of the power of science in general, and physics in particular. The underlying connections and unity of physics allow certain great minds to solve puzzles without having all the pieces. The prediction of electromagnetic waves is one of the most spectacular examples of this power. Certain others, such as the prediction of antimatter, will be discussed in later modules.

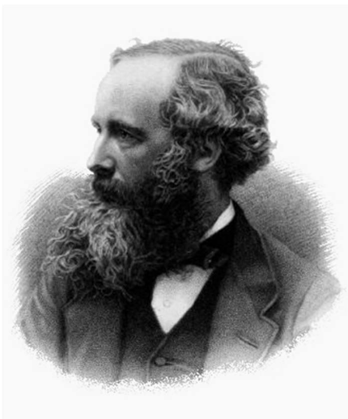
The electromagnetic waves sent and received by this 50-foot radar dish antenna at Kennedy Space Center in Florida are not visible, but help track expendable launch vehicles with high-definition imagery. The first use of this C-band radar dish was for the launch of the Atlas V rocket sending the New Horizons probe

toward Pluto.
(credit: NASA)

Maxwell’s Equations: Electromagnetic Waves Predicted and Observed

- Restate Maxwell’s equations.

The Scotsman James Clerk Maxwell (1831–1879) is regarded as the greatest theoretical physicist of the 19th century. (See [link].) Although he died young, Maxwell not only formulated a complete electromagnetic theory, represented by **Maxwell’s equations**, he also developed the kinetic theory of gases and made significant contributions to the understanding of color vision and the nature of Saturn’s rings.



James Clerk Maxwell, a 19th-century physicist, developed a theory that explained the relationship between electricity and magnetism and correctly predicted that visible light is caused by electromagnetic

waves. (credit: G. J. Stodart)

Maxwell brought together all the work that had been done by brilliant physicists such as Oersted, Coulomb, Gauss, and Faraday, and added his own insights to develop the overarching theory of electromagnetism. Maxwell’s equations are paraphrased here in words because their mathematical statement is beyond the level of this text. However, the equations illustrate how apparently simple mathematical statements can elegantly unite and express a multitude of concepts—why mathematics is the language of science.

Note:
Maxwell’s Equations

1. **Electric field lines** originate on positive charges and terminate on negative charges. The electric field is defined as the force per unit charge on a test charge, and the strength of the force is related to the electric constant ε_0, also known as the permittivity of free space. From Maxwell’s first equation we obtain a special form of Coulomb’s law known as Gauss’s law for electricity.
2. **Magnetic field lines** are continuous, having no beginning or end. No magnetic monopoles are known to exist. The strength of the magnetic force is related to the magnetic constant μ_0, also known as the permeability of free space. This second of Maxwell’s equations is known as Gauss’s law for magnetism.
3. A changing magnetic field induces an electromotive force (emf) and, hence, an electric field. The direction of the emf opposes the change. This third of Maxwell’s equations is Faraday’s law of induction, and includes Lenz’s law.
4. Magnetic fields are generated by moving charges or by changing electric fields. This fourth of Maxwell’s equations encompasses Ampere’s law and adds another source of magnetism—changing electric fields.

Maxwell's equations encompass the major laws of electricity and magnetism. What is not so apparent is the symmetry that Maxwell introduced in his mathematical framework. Especially important is his addition of the hypothesis that changing electric fields create magnetic fields. This is exactly analogous (and symmetric) to Faraday's law of induction and had been suspected for some time, but fits beautifully into Maxwell's equations.

Symmetry is apparent in nature in a wide range of situations. In contemporary research, symmetry plays a major part in the search for sub-atomic particles using massive multinational particle accelerators such as the new Large Hadron Collider at CERN.

Note:
Making Connections: Unification of Forces
Maxwell's complete and symmetric theory showed that electric and magnetic forces are not separate, but different manifestations of the same thing—the electromagnetic force. This classical unification of forces is one motivation for current attempts to unify the four basic forces in nature—the gravitational, electrical, strong, and weak nuclear forces.

Since changing electric fields create relatively weak magnetic fields, they could not be easily detected at the time of Maxwell's hypothesis. Maxwell realized, however, that oscillating charges, like those in AC circuits, produce changing electric fields. He predicted that these changing fields would propagate from the source like waves generated on a lake by a jumping fish.

The waves predicted by Maxwell would consist of oscillating electric and magnetic fields—defined to be an electromagnetic wave (EM wave). Electromagnetic waves would be capable of exerting forces on charges great distances from their source, and they might thus be detectable. Maxwell calculated that electromagnetic waves would propagate at a speed given by the equation

Equation:

$$c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}}.$$

When the values for μ_0 and ε_0 are entered into the equation for c, we find that

Equation:

$$c = \frac{1}{\sqrt{\left(8.85 \times 10^{-12}\ \frac{\mathrm{C}^2}{\mathrm{N{\cdot}m}^2}\right)\left(4\pi \times 10^{-7}\ \frac{\mathrm{T{\cdot}m}}{\mathrm{A}}\right)}} = 3.00 \times 10^8\ \mathrm{m/s},$$

which is the speed of light. In fact, Maxwell concluded that light is an electromagnetic wave having such wavelengths that it can be detected by the eye.

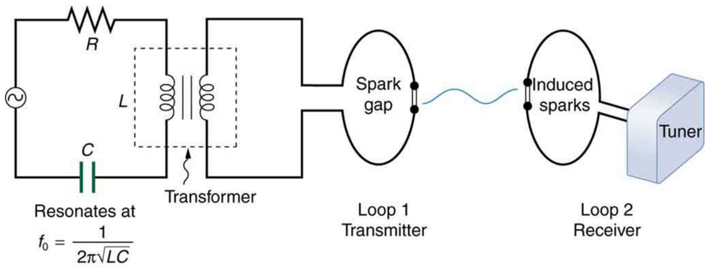
Other wavelengths should exist—it remained to be seen if they did. If so, Maxwell's theory and remarkable predictions would be verified, the greatest triumph of physics since Newton. Experimental verification came within a few years, but not before Maxwell's death.

Hertz's Observations

The German physicist Heinrich Hertz (1857–1894) was the first to generate and detect certain types of electromagnetic waves in the laboratory. Starting in 1887, he performed a series of experiments that not only confirmed the existence of electromagnetic waves, but also verified that they travel at the speed of light.

Hertz used an AC RLC (resistor-inductor-capacitor) circuit that resonates at a known frequency $f_0 = \frac{1}{2\pi\sqrt{\mathrm{LC}}}$ and connected it to a loop of wire as shown in [link]. High voltages induced across the gap in the loop produced sparks that were visible evidence of the current in the circuit and that helped generate electromagnetic waves.

Across the laboratory, Hertz had another loop attached to another RLC circuit, which could be tuned (as the dial on a radio) to the same resonant frequency as the first and could, thus, be made to receive electromagnetic waves. This loop also had a gap across which sparks were generated, giving solid evidence that electromagnetic waves had been received.



The apparatus used by Hertz in 1887 to generate and detect electromagnetic waves. An RLC circuit connected to the first loop caused sparks across a gap in the wire loop and generated electromagnetic waves. Sparks across a gap in the second loop located across the laboratory gave evidence that the waves had been received.

Hertz also studied the reflection, refraction, and interference patterns of the electromagnetic waves he generated, verifying their wave character. He was able to determine wavelength from the interference patterns, and knowing their frequency, he could calculate the propagation speed using the equation $v = f\lambda$ (velocity—or speed—equals frequency times wavelength). Hertz was thus able to prove that electromagnetic waves travel at the speed of light. The SI unit for frequency, the hertz ($1\ \mathrm{Hz} = 1\ \mathrm{cycle/sec}$), is named in his honor.

Section Summary

- Electromagnetic waves consist of oscillating electric and magnetic fields and propagate at the speed of light c. They were predicted by Maxwell, who also showed that
  Equation:

  $$c=\frac{1}{\sqrt{\mu_0\varepsilon_0}},$$

  where μ_0 is the permeability of free space and ε_0 is the permittivity of free space.
- Maxwell's prediction of electromagnetic waves resulted from his formulation of a complete and symmetric theory of electricity and magnetism, known as Maxwell's equations.
- These four equations are paraphrased in this text, rather than presented numerically, and encompass the major laws of electricity and magnetism. First is Gauss's law for electricity, second is Gauss's law for magnetism, third is Faraday's law of induction, including Lenz's law, and fourth is Ampere's law in a symmetric formulation that adds another source of magnetism—changing electric fields.

Problems & Exercises

Exercise:

Problem:

Verify that the correct value for the speed of light c is obtained when numerical values for the permeability and permittivity of free space (μ_0 and ε_0) are entered into the equation $c=\frac{1}{\sqrt{\mu_0\varepsilon_0}}$.

Exercise:

Problem:

Show that, when SI units for μ_0 and ε_0 are entered, the units given by the right-hand side of the equation in the problem above are m/s.

Glossary

electromagnetic waves

radiation in the form of waves of electric and magnetic energy

Maxwell's equations

a set of four equations that comprise a complete, overarching theory of electromagnetism

RLC circuit

an electric circuit that includes a resistor, capacitor and inductor

hertz

an SI unit denoting the frequency of an electromagnetic wave, in cycles per second

speed of light

in a vacuum, such as space, the speed of light is a constant 3×10^8 m/s

electromotive force (emf)

energy produced per unit charge, drawn from a source that produces an electrical current

electric field lines

a pattern of imaginary lines that extend between an electric source and charged objects in the surrounding area, with arrows pointed away from positively charged objects and toward negatively charged objects. The more lines in the pattern, the stronger the electric field in that region

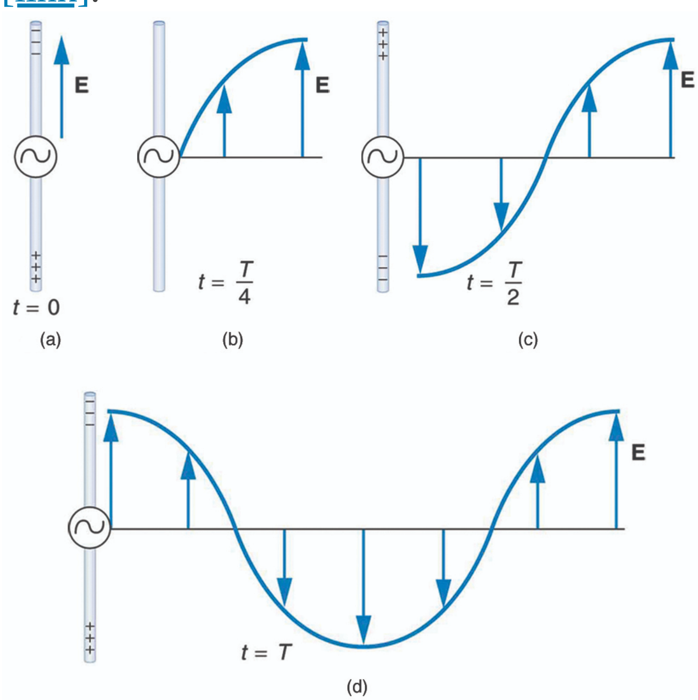
magnetic field lines

a pattern of continuous, imaginary lines that emerge from and enter into opposite magnetic poles. The density of the lines indicates the magnitude of the magnetic field

Production of Electromagnetic Waves

- Describe the electric and magnetic waves as they move out from a source, such as an AC generator.
- Explain the mathematical relationship between the magnetic field strength and the electrical field strength.
- Calculate the maximum strength of the magnetic field in an electromagnetic wave, given the maximum electric field strength.

We can get a good understanding of **electromagnetic waves** (EM) by considering how they are produced. Whenever a current varies, associated electric and magnetic fields vary, moving out from the source like waves. Perhaps the easiest situation to visualize is a varying current in a long straight wire, produced by an AC generator at its center, as illustrated in [link].



This long straight gray wire with an AC generator at its center becomes a broadcast antenna for electromagnetic waves. Shown here are the charge distributions at four different times. The electric field ($\mathbf{E}$) propagates away

from the antenna at the speed of light, forming part of an electromagnetic wave.

The **electric field** ($\mathbf{E}$) shown surrounding the wire is produced by the charge distribution on the wire. Both the $\mathbf{E}$ and the charge distribution vary as the current changes. The changing field propagates outward at the speed of light.

There is an associated **magnetic field** ($\mathbf{B}$) which propagates outward as well (see [link]). The electric and magnetic fields are closely related and propagate as an electromagnetic wave. This is what happens in broadcast antennae such as those in radio and TV stations.

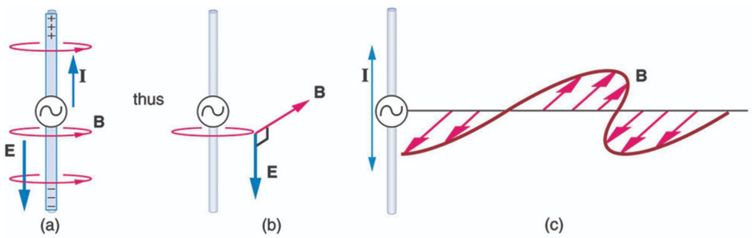
Closer examination of the one complete cycle shown in [link] reveals the periodic nature of the generator-driven charges oscillating up and down in the antenna and the electric field produced. At time $t = 0$, there is the maximum separation of charge, with negative charges at the top and positive charges at the bottom, producing the maximum magnitude of the electric field (or E-field) in the upward direction. One-fourth of a cycle later, there is no charge separation and the field next to the antenna is zero, while the maximum E-field has moved away at speed c.

As the process continues, the charge separation reverses and the field reaches its maximum downward value, returns to zero, and rises to its maximum upward value at the end of one complete cycle. The outgoing wave has an **amplitude** proportional to the maximum separation of charge. Its **wavelength**(λ) is proportional to the period of the oscillation and, hence, is smaller for short periods or high frequencies. (As usual, wavelength and **frequency**(f) are inversely proportional.)

Electric and Magnetic Waves: Moving Together

Following Ampere's law, current in the antenna produces a magnetic field, as shown in [link]. The relationship between $\mathbf{E}$ and $\mathbf{B}$ is shown at one

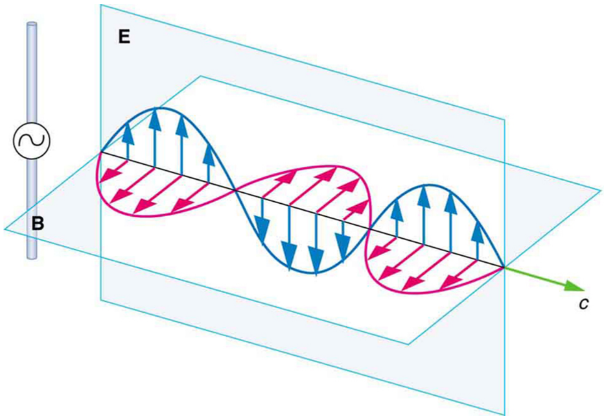
instant in [link] (a). As the current varies, the magnetic field varies in magnitude and direction.



(a) The current in the antenna produces the circular magnetic field lines. The current ($\mathbf{I}$) produces the separation of charge along the wire, which in turn creates the electric field as shown. (b) The electric and magnetic fields ($\mathbf{E}$ and $\mathbf{B}$) near the wire are perpendicular; they are shown here for one point in space. (c) The magnetic field varies with current and propagates away from the antenna at the speed of light.

The magnetic field lines also propagate away from the antenna at the speed of light, forming the other part of the electromagnetic wave, as seen in [link] (b). The magnetic part of the wave has the same period and wavelength as the electric part, since they are both produced by the same movement and separation of charges in the antenna.

The electric and magnetic waves are shown together at one instant in time in [link]. The electric and magnetic fields produced by a long straight wire antenna are exactly in phase. Note that they are perpendicular to one another and to the direction of propagation, making this a **transverse wave**.



A part of the electromagnetic wave sent out from the antenna at one instant in time. The electric and magnetic fields ($\mathbf{E}$ and $\mathbf{B}$) are in phase, and they are perpendicular to one another and the direction of propagation. For clarity, the waves are shown only along one direction, but they propagate out in other directions too.

Electromagnetic waves generally propagate out from a source in all directions, sometimes forming a complex radiation pattern. A linear antenna like this one will not radiate parallel to its length, for example. The wave is shown in one direction from the antenna in [link] to illustrate its basic characteristics.

Instead of the AC generator, the antenna can also be driven by an AC circuit. In fact, charges radiate whenever they are accelerated. But while a current in a circuit needs a complete path, an antenna has a varying charge distribution forming a **standing wave**, driven by the AC. The dimensions of the antenna are critical for determining the frequency of the radiated electromagnetic waves. This is a **resonant** phenomenon and when we tune

radios or TV, we vary electrical properties to achieve appropriate resonant conditions in the antenna.

Receiving Electromagnetic Waves

Electromagnetic waves carry energy away from their source, similar to a sound wave carrying energy away from a standing wave on a guitar string. An antenna for receiving EM signals works in reverse. And like antennas that produce EM waves, receiver antennas are specially designed to resonate at particular frequencies.

An incoming electromagnetic wave accelerates electrons in the antenna, setting up a standing wave. If the radio or TV is switched on, electrical components pick up and amplify the signal formed by the accelerating electrons. The signal is then converted to audio and/or video format. Sometimes big receiver dishes are used to focus the signal onto an antenna.

In fact, charges radiate whenever they are accelerated. When designing circuits, we often assume that energy does not quickly escape AC circuits, and mostly this is true. A broadcast antenna is specially designed to enhance the rate of electromagnetic radiation, and shielding is necessary to keep the radiation close to zero. Some familiar phenomena are based on the production of electromagnetic waves by varying currents. Your microwave oven, for example, sends electromagnetic waves, called microwaves, from a concealed antenna that has an oscillating current imposed on it.

Relating E-Field and B-Field Strengths

There is a relationship between the E- and B-field strengths in an electromagnetic wave. This can be understood by again considering the antenna just described. The stronger the E-field created by a separation of charge, the greater the current and, hence, the greater the B-field created.

Since current is directly proportional to voltage (Ohm’s law) and voltage is directly proportional to E-field strength, the two should be directly proportional. It can be shown that the magnitudes of the fields do have a constant ratio, equal to the speed of light. That is,

Equation:

$$\frac{E}{B} = c$$

is the ratio of E-field strength to B-field strength in any electromagnetic wave. This is true at all times and at all locations in space. A simple and elegant result.

Example:
Calculating B-Field Strength in an Electromagnetic Wave
What is the maximum strength of the B-field in an electromagnetic wave that has a maximum E-field strength of $1000\ \text{V/m}$?
Strategy
To find the B-field strength, we rearrange the above equation to solve for B, yielding
Equation:

$$B = \frac{E}{c}.$$

Solution
We are given E, and c is the speed of light. Entering these into the expression for B yields
Equation:

$$B = \frac{1000\ \text{V/m}}{3.00 \times 10^{8}\ \text{m/s}} = 3.33 \times 10^{-6}\ \text{T},$$

Where T stands for Tesla, a measure of magnetic field strength.
Discussion
The B-field strength is less than a tenth of the Earth's admittedly weak magnetic field. This means that a relatively strong electric field of 1000 V/m is accompanied by a relatively weak magnetic field. Note that as this

wave spreads out, say with distance from an antenna, its field strengths become progressively weaker.

The result of this example is consistent with the statement made in the module Maxwell’s Equations: Electromagnetic Waves Predicted and Observed that changing electric fields create relatively weak magnetic fields. They can be detected in electromagnetic waves, however, by taking advantage of the phenomenon of resonance, as Hertz did. A system with the same natural frequency as the electromagnetic wave can be made to oscillate. All radio and TV receivers use this principle to pick up and then amplify weak electromagnetic waves, while rejecting all others not at their resonant frequency.

Note:
Take-Home Experiment: Antennas
For your TV or radio at home, identify the antenna, and sketch its shape. If you don’t have cable, you might have an outdoor or indoor TV antenna. Estimate its size. If the TV signal is between 60 and 216 MHz for basic channels, then what is the wavelength of those EM waves?
Try tuning the radio and note the small range of frequencies at which a reasonable signal for that station is received. (This is easier with digital readout.) If you have a car with a radio and extendable antenna, note the quality of reception as the length of the antenna is changed.

Note:
PhET Explorations: Radio Waves and Electromagnetic Fields
Broadcast radio waves from KPhET. Wiggle the transmitter electron manually or have it oscillate automatically. Display the field as a curve or vectors. The strip chart shows the electron positions at the transmitter and at the receiver.
https://archive.cnx.org/specials/c8dd764c-ae74-11e5-af4c-3375261fa183/radio-waves/#sim-radio-waves

Section Summary

- Electromagnetic waves are created by oscillating charges (which radiate whenever accelerated) and have the same frequency as the oscillation.
- Since the electric and magnetic fields in most electromagnetic waves are perpendicular to the direction in which the wave moves, it is ordinarily a transverse wave.
- The strengths of the electric and magnetic parts of the wave are related by

  Equation:

  $$\frac{E}{B} = \text{c},$$

  which implies that the magnetic field B is very weak relative to the electric field E.

Conceptual Questions

Exercise:

Problem:

The direction of the electric field shown in each part of [link] is that produced by the charge distribution in the wire. Justify the direction shown in each part, using the Coulomb force law and the definition of $\mathbf{E} = \mathbf{F}/q$, where q is a positive test charge.

Exercise:

Problem:

Is the direction of the magnetic field shown in [link] (a) consistent with the right-hand rule for current (RHR-2) in the direction shown in the figure?

Exercise:

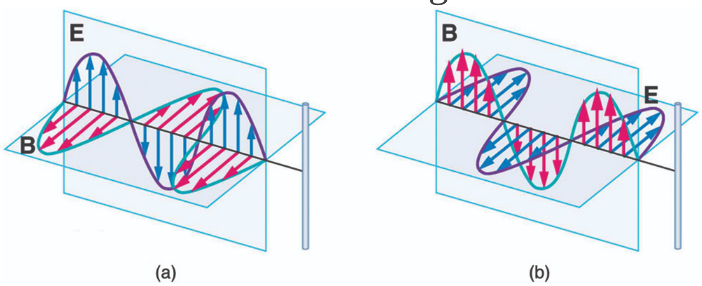
Problem:

Why is the direction of the current shown in each part of [link] opposite to the electric field produced by the wire's charge separation?

Exercise:

Problem:

In which situation shown in [link] will the electromagnetic wave be more successful in inducing a current in the wire? Explain.

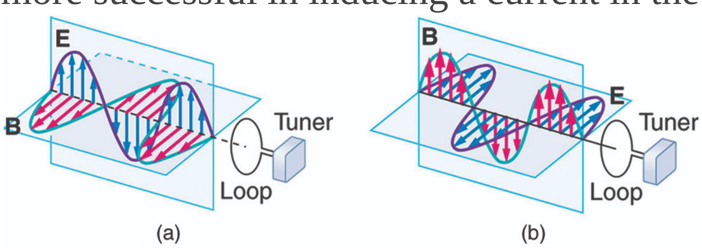


Electromagnetic waves approaching long straight wires.

Exercise:

Problem:

In which situation shown in [link] will the electromagnetic wave be more successful in inducing a current in the loop? Explain.



Electromagnetic waves approaching a wire loop.

Exercise:

Problem:

Should the straight wire antenna of a radio be vertical or horizontal to best receive radio waves broadcast by a vertical transmitter antenna? How should a loop antenna be aligned to best receive the signals? (Note that the direction of the loop that produces the best reception can be used to determine the location of the source. It is used for that purpose in tracking tagged animals in nature studies, for example.)

Exercise:

Problem:

Under what conditions might wires in a DC circuit emit electromagnetic waves?

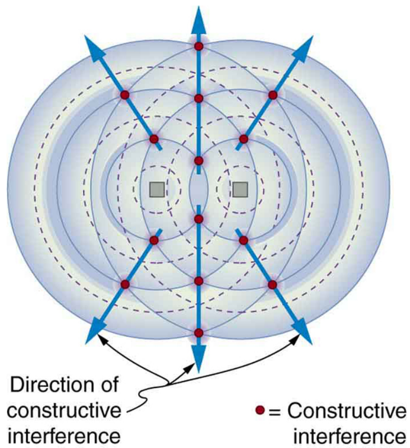
Exercise:

Problem: Give an example of interference of electromagnetic waves.

Exercise:

Problem:

[link] shows the interference pattern of two radio antennas broadcasting the same signal. Explain how this is analogous to the interference pattern for sound produced by two speakers. Could this be used to make a directional antenna system that broadcasts preferentially in certain directions? Explain.



An overhead view of two radio broadcast antennas sending the same signal, and the interference pattern they produce.

Exercise:

Problem: Can an antenna be any length? Explain your answer.

Problems & Exercises

Exercise:

Problem:

What is the maximum electric field strength in an electromagnetic wave that has a maximum magnetic field strength of $5.00 \times 10^{-4}\ \mathrm{T}$ (about 10 times the Earth’s)?

Solution:

150 kV/m

Exercise:

Problem:

The maximum magnetic field strength of an electromagnetic field is $5 \times 10^{-6}\ \text{T}$. Calculate the maximum electric field strength if the wave is traveling in a medium in which the speed of the wave is $0.75c$.

Exercise:

Problem:

Verify the units obtained for magnetic field strength B in [link] (using the equation $B = \frac{E}{c}$) are in fact teslas (T).

Glossary

electric field
: a vector quantity (**E**); the lines of electric force per unit charge, moving radially outward from a positive charge and in toward a negative charge

electric field strength
: the magnitude of the electric field, denoted *E*-field

magnetic field
: a vector quantity (**B**); can be used to determine the magnetic force on a moving charged particle

magnetic field strength
: the magnitude of the magnetic field, denoted *B*-field

transverse wave
: a wave, such as an electromagnetic wave, which oscillates perpendicular to the axis along the line of travel

standing wave

a wave that oscillates in place, with nodes where no motion happens

wavelength
: the distance from one peak to the next in a wave

amplitude
: the height, or magnitude, of an electromagnetic wave

frequency
: the number of complete wave cycles (up-down-up) passing a given point within one second (cycles/second)

resonant
: a system that displays enhanced oscillation when subjected to a periodic disturbance of the same frequency as its natural frequency

oscillate
: to fluctuate back and forth in a steady beat

The Electromagnetic Spectrum

- List three “rules of thumb” that apply to the different frequencies along the electromagnetic spectrum.
- Explain why the higher the frequency, the shorter the wavelength of an electromagnetic wave.
- Draw a simplified electromagnetic spectrum, indicating the relative positions, frequencies, and spacing of the different types of radiation bands.
- List and explain the different methods by which electromagnetic waves are produced across the spectrum.

In this module we examine how electromagnetic waves are classified into categories such as radio, infrared, ultraviolet, and so on, so that we can understand some of their similarities as well as some of their differences. We will also find that there are many connections with previously discussed topics, such as wavelength and resonance. A brief overview of the production and utilization of electromagnetic waves is found in [link].

Type of EM wave	**Production**	**Applications**	**Life sciences aspect**	**Issues**
Radio & TV	Accelerating charges	Communications Remote controls	MRI	Requires controls for band use
Microwaves	Accelerating charges & thermal agitation	Communications Ovens Radar	Deep heating	Cell phone use
Infrared	Thermal agitations & electronic transitions	Thermal imaging Heating	Absorbed by atmosphere	Greenhouse effect
Visible light	Thermal agitations & electronic transitions	All pervasive	Photosynthesis Human vision	

Type of EM wave	Production	Applications	Life sciences aspect	Issues
Ultraviolet	Thermal agitations & electronic transitions	Sterilization Cancer control	Vitamin D production	Ozone depletion Cancer causing
X-rays	Inner electronic transitions and fast collisions	Medical Security	Medical diagnosis Cancer therapy	Cancer causing
Gamma rays	Nuclear decay	Nuclear medicineSecurity	Medical diagnosis Cancer therapy	Cancer causing Radiation damage

Electromagnetic Waves

Note:
Connections: Waves
There are many types of waves, such as water waves and even earthquakes. Among the many shared attributes of waves are propagation speed, frequency, and wavelength. These are always related by the expression $v_{\text{W}} = f\lambda$. This module concentrates on EM waves, but other modules contain examples of all of these characteristics for sound waves and submicroscopic particles.

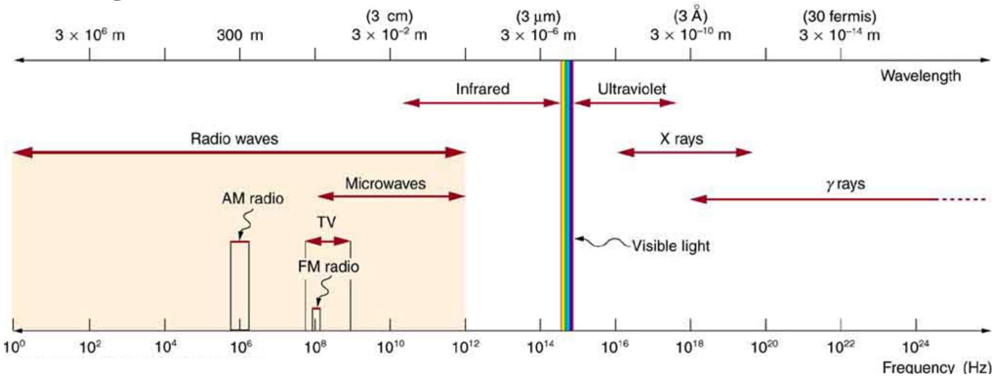
As noted before, an electromagnetic wave has a frequency and a wavelength associated with it and travels at the speed of light, or c. The relationship among these wave characteristics can be described by $v_{\text{W}} = f\lambda$, where v_{W} is the propagation speed of the wave, f is the frequency, and λ is the wavelength. Here $v_{\text{W}} = c$, so that for all electromagnetic waves,
Equation:

$$c = f\lambda.$$

Thus, for all electromagnetic waves, the greater the frequency, the smaller the wavelength.

[link] shows how the various types of electromagnetic waves are categorized according to their wavelengths and frequencies—that is, it shows the electromagnetic spectrum. Many of the

characteristics of the various types of electromagnetic waves are related to their frequencies and wavelengths, as we shall see.



The electromagnetic spectrum, showing the major categories of electromagnetic waves. The range of frequencies and wavelengths is remarkable. The dividing line between some categories is distinct, whereas other categories overlap.

Note:
Electromagnetic Spectrum: Rules of Thumb
Three rules that apply to electromagnetic waves in general are as follows:

- High-frequency electromagnetic waves are more energetic and are more able to penetrate than low-frequency waves.
- High-frequency electromagnetic waves can carry more information per unit time than low-frequency waves.
- The shorter the wavelength of any electromagnetic wave probing a material, the smaller the detail it is possible to resolve.

Note that there are exceptions to these rules of thumb.

Transmission, Reflection, and Absorption

What happens when an electromagnetic wave impinges on a material? If the material is transparent to the particular frequency, then the wave can largely be transmitted. If the material is opaque to the frequency, then the wave can be totally reflected. The wave can also be absorbed by the material, indicating that there is some interaction between the wave and the material, such as the thermal agitation of molecules.

Of course it is possible to have partial transmission, reflection, and absorption. We normally associate these properties with visible light, but they do apply to all electromagnetic waves.

What is not obvious is that something that is transparent to light may be opaque at other frequencies. For example, ordinary glass is transparent to visible light but largely opaque to ultraviolet radiation. Human skin is opaque to visible light—we cannot see through people—but transparent to X-rays.

Radio and TV Waves

The broad category of **radio waves** is defined to contain any electromagnetic wave produced by currents in wires and circuits. Its name derives from their most common use as a carrier of audio information (i.e., radio). The name is applied to electromagnetic waves of similar frequencies regardless of source. Radio waves from outer space, for example, do not come from alien radio stations. They are created by many astronomical phenomena, and their study has revealed much about nature on the largest scales.

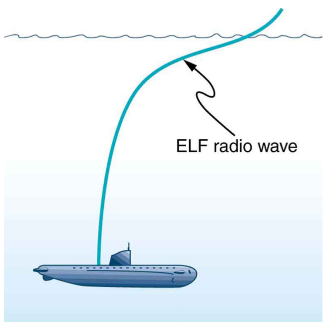
There are many uses for radio waves, and so the category is divided into many subcategories, including microwaves and those electromagnetic waves used for AM and FM radio, cellular telephones, and TV.

The lowest commonly encountered radio frequencies are produced by high-voltage AC power transmission lines at frequencies of 50 or 60 Hz. (See [link].) These extremely long wavelength electromagnetic waves (about 6000 km!) are one means of energy loss in long-distance power transmission.

This high-voltage traction power line running to Eutingen Railway Substation in Germany radiates electromagnetic waves with very long wavelengths. (credit: Zonk43, Wikimedia Commons)

There is an ongoing controversy regarding potential health hazards associated with exposure to these electromagnetic fields (E-fields). Some people suspect that living near such transmission lines may cause a variety of illnesses, including cancer. But demographic data are either inconclusive or simply do not support the hazard theory. Recent reports that have looked at many European and American epidemiological studies have found no increase in risk for cancer due to exposure to E-fields.

Extremely low frequency (ELF) radio waves of about 1 kHz are used to communicate with submerged submarines. The ability of radio waves to penetrate salt water is related to their wavelength (much like ultrasound penetrating tissue)—the longer the wavelength, the farther they penetrate. Since salt water is a good conductor, radio waves are strongly absorbed by it, and very long wavelengths are needed to reach a submarine under the surface. (See [link].)

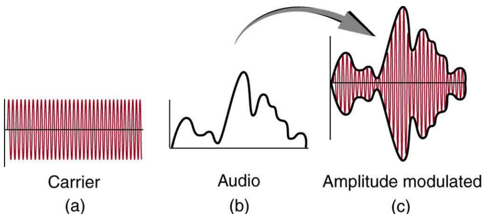


Very long wavelength radio waves are needed to reach this submarine, requiring extremely low frequency signals (ELF). Shorter wavelengths do not penetrate to any significant depth.

AM radio waves are used to carry commercial radio signals in the frequency range from 540 to 1600 kHz. The abbreviation AM stands for **amplitude modulation**, which is the method for placing information on these waves. (See [link].) A **carrier wave** having the basic frequency of the radio station, say 1530 kHz, is varied or modulated in amplitude by an audio signal. The resulting wave has a constant frequency, but a varying amplitude.

A radio receiver tuned to have the same resonant frequency as the carrier wave can pick up the signal, while rejecting the many other frequencies impinging on its antenna. The receiver’s

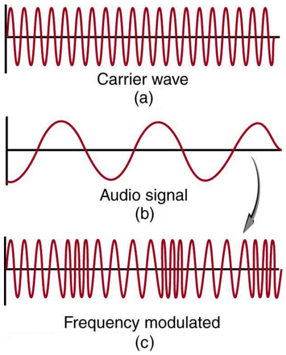
circuitry is designed to respond to variations in amplitude of the carrier wave to replicate the original audio signal. That audio signal is amplified to drive a speaker or perhaps to be recorded.



Amplitude modulation for AM radio. (a) A carrier wave at the station's basic frequency. (b) An audio signal at much lower audible frequencies. (c) The amplitude of the carrier is modulated by the audio signal without changing its basic frequency.

FM Radio Waves

FM radio waves are also used for commercial radio transmission, but in the frequency range of 88 to 108 MHz. FM stands for **frequency modulation**, another method of carrying information. (See [link].) Here a carrier wave having the basic frequency of the radio station, perhaps 105.1 MHz, is modulated in frequency by the audio signal, producing a wave of constant amplitude but varying frequency.



Frequency modulation for

FM radio. (a) A carrier wave at the station’s basic frequency. (b) An audio signal at much lower audible frequencies. (c) The frequency of the carrier is modulated by the audio signal without changing its amplitude.

Since audible frequencies range up to 20 kHz (or 0.020 MHz) at most, the frequency of the FM radio wave can vary from the carrier by as much as 0.020 MHz. Thus the carrier frequencies of two different radio stations cannot be closer than 0.020 MHz. An FM receiver is tuned to resonate at the carrier frequency and has circuitry that responds to variations in frequency, reproducing the audio information.

FM radio is inherently less subject to noise from stray radio sources than AM radio. The reason is that amplitudes of waves add. So an AM receiver would interpret noise added onto the amplitude of its carrier wave as part of the information. An FM receiver can be made to reject amplitudes other than that of the basic carrier wave and only look for variations in frequency. It is thus easier to reject noise from FM, since noise produces a variation in amplitude.

Television is also broadcast on electromagnetic waves. Since the waves must carry a great deal of visual as well as audio information, each channel requires a larger range of frequencies than simple radio transmission. TV channels utilize frequencies in the range of 54 to 88 MHz and 174 to 222 MHz. (The entire FM radio band lies between channels 88 MHz and 174 MHz.) These TV channels are called VHF (for **very high frequency**). Other channels called UHF (for **ultra high frequency**) utilize an even higher frequency range of 470 to 1000 MHz.

The TV video signal is AM, while the TV audio is FM. Note that these frequencies are those of free transmission with the user utilizing an old-fashioned roof antenna. Satellite dishes and cable transmission of TV occurs at significantly higher frequencies and is rapidly evolving with the use of the high-definition or HD format.

Example:
Calculating Wavelengths of Radio Waves

Calculate the wavelengths of a 1530-kHz AM radio signal, a 105.1-MHz FM radio signal, and a 1.90-GHz cell phone signal.

Strategy

The relationship between wavelength and frequency is $c = f\lambda$, where $c = 3.00 \times 10^8\ \text{m/s}$ is the speed of light (the speed of light is only very slightly smaller in air than it is in a vacuum). We can rearrange this equation to find the wavelength for all three frequencies.

Solution

Rearranging gives

Equation:

$$\lambda = \frac{c}{f}.$$

(a) For the $f = 1530\ \text{kHz}$ AM radio signal, then,

Equation:

$$\begin{aligned} \lambda &= \frac{3.00 \times 10^8\ \text{m/s}}{1530 \times 10^3\ \text{cycles/s}} \\ &= 196\ \text{m}. \end{aligned}$$

(b) For the $f = 105.1\ \text{MHz}$ FM radio signal,

Equation:

$$\begin{aligned} \lambda &= \frac{3.00 \times 10^8\ \text{m/s}}{105.1 \times 10^6\ \text{cycles/s}} \\ &= 2.85\ \text{m}. \end{aligned}$$

(c) And for the $f = 1.90\ \text{GHz}$ cell phone,

Equation:

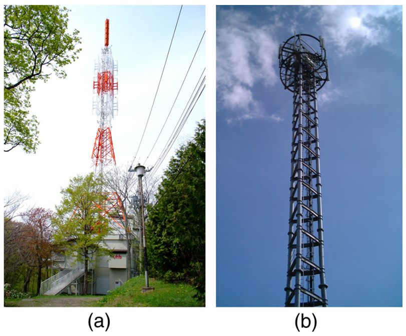
$$\begin{aligned} \lambda &= \frac{3.00 \times 10^8\ \text{m/s}}{1.90 \times 10^9\ \text{cycles/s}} \\ &= 0.158\ \text{m}. \end{aligned}$$

Discussion

These wavelengths are consistent with the spectrum in [link]. The wavelengths are also related to other properties of these electromagnetic waves, as we shall see.

The wavelengths found in the preceding example are representative of AM, FM, and cell phones, and account for some of the differences in how they are broadcast and how well they travel. The most efficient length for a linear antenna, such as discussed in Production of Electromagnetic Waves, is $\lambda/2$, half the wavelength of the electromagnetic wave. Thus a very large antenna is needed to efficiently broadcast typical AM radio with its carrier wavelengths on the order of hundreds of meters.

One benefit to these long AM wavelengths is that they can go over and around rather large obstacles (like buildings and hills), just as ocean waves can go around large rocks. FM and TV are best received when there is a line of sight between the broadcast antenna and receiver, and they are often sent from very tall structures. FM, TV, and mobile phone antennas themselves are much smaller than those used for AM, but they are elevated to achieve an unobstructed line of sight. (See [link].)



(a) A large tower is used to broadcast TV signals. The actual antennas are small structures on top of the tower—they are placed at great heights to have a clear line of sight over a large broadcast area. (credit: Ozizo, Wikimedia Commons) (b) The NTT Dokomo mobile phone tower at Tokorozawa City, Japan. (credit: tokoroten, Wikimedia Commons)

Radio Wave Interference

Astronomers and astrophysicists collect signals from outer space using electromagnetic waves. A common problem for astrophysicists is the “pollution” from electromagnetic radiation pervading our surroundings from communication systems in general. Even everyday gadgets like our car keys having the facility to lock car doors remotely and being able to turn TVs on and off using remotes involve radio-wave frequencies. In order to prevent interference between all these electromagnetic signals, strict regulations are drawn up for different organizations to utilize different radio frequency bands.

One reason why we are sometimes asked to switch off our mobile phones (operating in the range of 1.9 GHz) on airplanes and in hospitals is that important communications or medical equipment often uses similar radio frequencies and their operation can be affected by frequencies used in the communication devices.

For example, radio waves used in magnetic resonance imaging (MRI) have frequencies on the order of 100 MHz, although this varies significantly depending on the strength of the magnetic field used and the nuclear type being scanned. MRI is an important medical imaging and research tool, producing highly detailed two- and three-dimensional images. Radio waves are broadcast, absorbed, and reemitted in a resonance process that is sensitive to the density of nuclei (usually protons or hydrogen nuclei).

The wavelength of 100-MHz radio waves is 3 m, yet using the sensitivity of the resonant frequency to the magnetic field strength, details smaller than a millimeter can be imaged. This is a good example of an exception to a rule of thumb (in this case, the rubric that details much smaller than the probe's wavelength cannot be detected). The intensity of the radio waves used in MRI presents little or no hazard to human health.

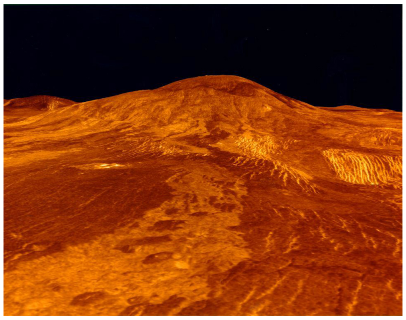
Microwaves

Microwaves are the highest-frequency electromagnetic waves that can be produced by currents in macroscopic circuits and devices. Microwave frequencies range from about $10^9\ \text{Hz}$ to the highest practical LC resonance at nearly $10^{12}\ \text{Hz}$. Since they have high frequencies, their wavelengths are short compared with those of other radio waves—hence the name "microwave."

Microwaves can also be produced by atoms and molecules. They are, for example, a component of electromagnetic radiation generated by **thermal agitation**. The thermal motion of atoms and molecules in any object at a temperature above absolute zero causes them to emit and absorb radiation.

Since it is possible to carry more information per unit time on high frequencies, microwaves are quite suitable for communications. Most satellite-transmitted information is carried on microwaves, as are land-based long-distance transmissions. A clear line of sight between transmitter and receiver is needed because of the short wavelengths involved.

Radar is a common application of microwaves that was first developed in World War II. By detecting and timing microwave echoes, radar systems can determine the distance to objects as diverse as clouds and aircraft. A Doppler shift in the radar echo can be used to determine the speed of a car or the intensity of a rainstorm. Sophisticated radar systems are used to map the Earth and other planets, with a resolution limited by wavelength. (See [link].) The shorter the wavelength of any probe, the smaller the detail it is possible to observe.



An image of Sif Mons with lava flows on Venus, based on Magellan synthetic aperture radar data combined with radar altimetry to produce a three-dimensional map of the surface. The Venusian atmosphere is opaque to visible light, but not to the microwaves that were used to create this image. (credit: NSSDC, NASA/JPL)

Heating with Microwaves

How does the ubiquitous microwave oven produce microwaves electronically, and why does food absorb them preferentially? Microwaves at a frequency of 2.45 GHz are produced by accelerating electrons. The microwaves are then used to induce an alternating electric field in the oven.

Water and some other constituents of food have a slightly negative charge at one end and a slightly positive charge at one end (called polar molecules). The range of microwave frequencies is specially selected so that the polar molecules, in trying to keep orienting themselves with the electric field, absorb these energies and increase their temperatures—called dielectric heating.

The energy thereby absorbed results in thermal agitation heating food and not the plate, which does not contain water. Hot spots in the food are related to constructive and destructive interference patterns. Rotating antennas and food turntables help spread out the hot spots.

Another use of microwaves for heating is within the human body. Microwaves will penetrate more than shorter wavelengths into tissue and so can accomplish “deep heating” (called

microwave diathermy). This is used for treating muscular pains, spasms, tendonitis, and rheumatoid arthritis.

Note:
Making Connections: Take-Home Experiment—Microwave Ovens

1. Look at the door of a microwave oven. Describe the structure of the door. Why is there a metal grid on the door? How does the size of the holes in the grid compare with the wavelengths of microwaves used in microwave ovens? What is this wavelength?
2. Place a glass of water (about 250 ml) in the microwave and heat it for 30 seconds. Measure the temperature gain (the $\Delta \mathrm{T}$). Assuming that the power output of the oven is 1000 W, calculate the efficiency of the heat-transfer process.
3. Remove the rotating turntable or moving plate and place a cup of water in several places along a line parallel with the opening. Heat for 30 seconds and measure the $\Delta \mathrm{T}$ for each position. Do you see cases of destructive interference?

Microwaves generated by atoms and molecules far away in time and space can be received and detected by electronic circuits. Deep space acts like a blackbody with a 2.7 K temperature, radiating most of its energy in the microwave frequency range. In 1964, Penzias and Wilson detected this radiation and eventually recognized that it was the radiation of the Big Bang's cooled remnants.

Infrared Radiation

The microwave and infrared regions of the electromagnetic spectrum overlap (see [link]). **Infrared radiation** is generally produced by thermal motion and the vibration and rotation of atoms and molecules. Electronic transitions in atoms and molecules can also produce infrared radiation.

The range of infrared frequencies extends up to the lower limit of visible light, just below red. In fact, infrared means "below red." Frequencies at its upper limit are too high to be produced by accelerating electrons in circuits, but small systems, such as atoms and molecules, can vibrate fast enough to produce these waves.

Water molecules rotate and vibrate particularly well at infrared frequencies, emitting and absorbing them so efficiently that the emissivity for skin is $e = 0.97$ in the infrared. Night-vision scopes can detect the infrared emitted by various warm objects, including humans, and convert it to visible light.

We can examine radiant heat transfer from a house by using a camera capable of detecting infrared radiation. Reconnaissance satellites can detect buildings, vehicles, and even individual humans by their infrared emissions, whose power radiation is proportional to the fourth power of the absolute temperature. More mundanely, we use infrared lamps, some of which are called

quartz heaters, to preferentially warm us because we absorb infrared better than our surroundings.

The Sun radiates like a nearly perfect blackbody (that is, it has $e = 1$), with a 6000 K surface temperature. About half of the solar energy arriving at the Earth is in the infrared region, with most of the rest in the visible part of the spectrum, and a relatively small amount in the ultraviolet. On average, 50 percent of the incident solar energy is absorbed by the Earth.

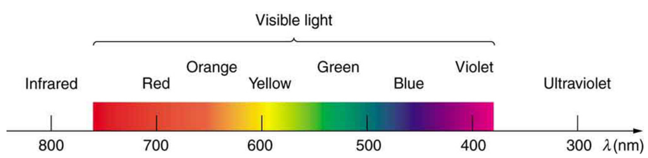
The relatively constant temperature of the Earth is a result of the energy balance between the incoming solar radiation and the energy radiated from the Earth. Most of the infrared radiation emitted from the Earth is absorbed by CO_2 and H_2O in the atmosphere and then radiated back to Earth or into outer space. This radiation back to Earth is known as the greenhouse effect, and it maintains the surface temperature of the Earth about $40^{\circ}C$ higher than it would be if there is no absorption. Some scientists think that the increased concentration of CO_2 and other greenhouse gases in the atmosphere, resulting from increases in fossil fuel burning, has increased global average temperatures.

Visible Light

Visible light is the narrow segment of the electromagnetic spectrum to which the normal human eye responds. Visible light is produced by vibrations and rotations of atoms and molecules, as well as by electronic transitions within atoms and molecules. The receivers or detectors of light largely utilize electronic transitions. We say the atoms and molecules are excited when they absorb and relax when they emit through electronic transitions.

[link] shows this part of the spectrum, together with the colors associated with particular pure wavelengths. We usually refer to visible light as having wavelengths of between 400 nm and 750 nm. (The retina of the eye actually responds to the lowest ultraviolet frequencies, but these do not normally reach the retina because they are absorbed by the cornea and lens of the eye.)

Red light has the lowest frequencies and longest wavelengths, while violet has the highest frequencies and shortest wavelengths. Blackbody radiation from the Sun peaks in the visible part of the spectrum but is more intense in the red than in the violet, making the Sun yellowish in appearance.



A small part of the electromagnetic spectrum that includes its visible components. The divisions between infrared, visible, and ultraviolet are not perfectly

distinct, nor are those between the seven rainbow colors.

Living things—plants and animals—have evolved to utilize and respond to parts of the electromagnetic spectrum they are embedded in. Visible light is the most predominant and we enjoy the beauty of nature through visible light. Plants are more selective. Photosynthesis makes use of parts of the visible spectrum to make sugars.

Example:
Integrated Concept Problem: Correcting Vision with Lasers
During laser vision correction, a brief burst of 193-nm ultraviolet light is projected onto the cornea of a patient. It makes a spot 0.80 mm in diameter and evaporates a layer of cornea $0.30\ \mu\text{m}$ thick. Calculate the energy absorbed, assuming the corneal tissue has the same properties as water; it is initially at 34^{o}C. Assume the evaporated tissue leaves at a temperature of 100^{o}C.
Strategy
The energy from the laser light goes toward raising the temperature of the tissue and also toward evaporating it. Thus we have two amounts of heat to add together. Also, we need to find the mass of corneal tissue involved.
Solution
To figure out the heat required to raise the temperature of the tissue to 100^{o}C, we can apply concepts of thermal energy. We know that
Equation:

$$\text{Q} = \text{mc}\Delta T,$$

where Q is the heat required to raise the temperature, ΔT is the desired change in temperature, m is the mass of tissue to be heated, and c is the specific heat of water equal to 4186 J/kg/K. Without knowing the mass m at this point, we have
Equation:

$$Q = m(4186\ \text{J/kg/K})(100^{\text{o}}\text{C}–34^{\text{o}}\text{C}) = m(276{,}276\ \text{J/kg}) = m(276\ \text{kJ/kg}\ ).$$

The latent heat of vaporization of water is 2256 kJ/kg, so that the energy needed to evaporate mass m is
Equation:

$$Q_{\text{v}} = mL_{\text{v}} = m(2256\ \text{kJ/kg}).$$

To find the mass m, we use the equation $\rho = m/\text{V}$, where ρ is the density of the tissue and V is its volume. For this case,
Equation:

$$\begin{aligned} m &= \rho \mathrm{V} \\ &= (1000\ \mathrm{kg/m^3})(\mathrm{area} \times \mathrm{thickness}(\mathrm{m^3})) \\ &= (1000\ \mathrm{kg/m^3})(\pi(0.80 \times 10^{-3}\ \mathrm{m})^2/4)(0.30 \times 10^{-6}\ \mathrm{m}) \\ &= 0.151 \times 10^{-9}\ \mathrm{kg}. \end{aligned}$$

Therefore, the total energy absorbed by the tissue in the eye is the sum of Q and $\mathrm{Q_v}$:
Equation:

$$\mathrm{Q_{tot}} = m(c\Delta\mathrm{T} + \mathrm{L_v}) = (0.151 \times 10^{-9}\ \mathrm{kg})(276\ \mathrm{kJ/kg} + 2256\ \mathrm{kJ/kg}) = 382 \times 10^{-9}\ \mathrm{kJ}.$$

Discussion
The lasers used for this eye surgery are excimer lasers, whose light is well absorbed by biological tissue. They evaporate rather than burn the tissue, and can be used for precision work. Most lasers used for this type of eye surgery have an average power rating of about one watt. For our example, if we assume that each laser burst from this pulsed laser lasts for 10 ns, and there are 400 bursts per second, then the average power is $\mathrm{Q_{tot}} \times 400 = 150\ \mathrm{mW}$.

Optics is the study of the behavior of visible light and other forms of electromagnetic waves. Optics falls into two distinct categories. When electromagnetic radiation, such as visible light, interacts with objects that are large compared with its wavelength, its motion can be represented by straight lines like rays. Ray optics is the study of such situations and includes lenses and mirrors.

When electromagnetic radiation interacts with objects about the same size as the wavelength or smaller, its wave nature becomes apparent. For example, observable detail is limited by the wavelength, and so visible light can never detect individual atoms, because they are so much smaller than its wavelength. Physical or wave optics is the study of such situations and includes all wave characteristics.

Note:
Take-Home Experiment: Colors That Match
When you light a match you see largely orange light; when you light a gas stove you see blue light. Why are the colors different? What other colors are present in these?

Ultraviolet Radiation

Ultraviolet means “above violet.” The electromagnetic frequencies of **ultraviolet radiation (UV)** extend upward from violet, the highest-frequency visible light. Ultraviolet is also produced by atomic and molecular motions and electronic transitions. The wavelengths of ultraviolet extend from 400 nm down to about 10 nm at its highest frequencies, which overlap

with the lowest X-ray frequencies. It was recognized as early as 1801 by Johann Ritter that the solar spectrum had an invisible component beyond the violet range.

Solar UV radiation is broadly subdivided into three regions: UV-A (320–400 nm), UV-B (290–320 nm), and UV-C (220–290 nm), ranked from long to shorter wavelengths (from smaller to larger energies). Most UV-B and all UV-C is absorbed by ozone (O_3) molecules in the upper atmosphere. Consequently, 99% of the solar UV radiation reaching the Earth's surface is UV-A.

Human Exposure to UV Radiation

It is largely exposure to UV-B that causes skin cancer. It is estimated that as many as 20% of adults will develop skin cancer over the course of their lifetime. Again, treatment is often successful if caught early. Despite very little UV-B reaching the Earth's surface, there are substantial increases in skin-cancer rates in countries such as Australia, indicating how important it is that UV-B and UV-C continue to be absorbed by the upper atmosphere.

All UV radiation can damage collagen fibers, resulting in an acceleration of the aging process of skin and the formation of wrinkles. Because there is so little UV-B and UV-C reaching the Earth's surface, sunburn is caused by large exposures, and skin cancer from repeated exposure. Some studies indicate a link between overexposure to the Sun when young and melanoma later in life.

The tanning response is a defense mechanism in which the body produces pigments to absorb future exposures in inert skin layers above living cells. Basically UV-B radiation excites DNA molecules, distorting the DNA helix, leading to mutations and the possible formation of cancerous cells.

Repeated exposure to UV-B may also lead to the formation of cataracts in the eyes—a cause of blindness among people living in the equatorial belt where medical treatment is limited. Cataracts, clouding in the eye's lens and a loss of vision, are age related; 60% of those between the ages of 65 and 74 will develop cataracts. However, treatment is easy and successful, as one replaces the lens of the eye with a plastic lens. Prevention is important. Eye protection from UV is more effective with plastic sunglasses than those made of glass.

A major acute effect of extreme UV exposure is the suppression of the immune system, both locally and throughout the body.

Low-intensity ultraviolet is used to sterilize haircutting implements, implying that the energy associated with ultraviolet is deposited in a manner different from lower-frequency electromagnetic waves. (Actually this is true for all electromagnetic waves with frequencies greater than visible light.)

Flash photography is generally not allowed of precious artworks and colored prints because the UV radiation from the flash can cause photo-degradation in the artworks. Often artworks will have an extra-thick layer of glass in front of them, which is especially designed to absorb UV radiation.

UV Light and the Ozone Layer

If all of the Sun's ultraviolet radiation reached the Earth's surface, there would be extremely grave effects on the biosphere from the severe cell damage it causes. However, the layer of ozone (O_3) in our upper atmosphere (10 to 50 km above the Earth) protects life by absorbing most of the dangerous UV radiation.

Unfortunately, today we are observing a depletion in ozone concentrations in the upper atmosphere. This depletion has led to the formation of an "ozone hole" in the upper atmosphere. The hole is more centered over the southern hemisphere, and changes with the seasons, being largest in the spring. This depletion is attributed to the breakdown of ozone molecules by refrigerant gases called chlorofluorocarbons (CFCs).

The UV radiation helps dissociate the CFC's, releasing highly reactive chlorine (Cl) atoms, which catalyze the destruction of the ozone layer. For example, the reaction of $\mathrm{CFCl_3}$ with a photon of light (hv) can be written as:

Equation:

$$\mathrm{CFCl_3} + \mathrm{h}v \rightarrow \mathrm{CFCl_2} + \mathrm{Cl.}$$

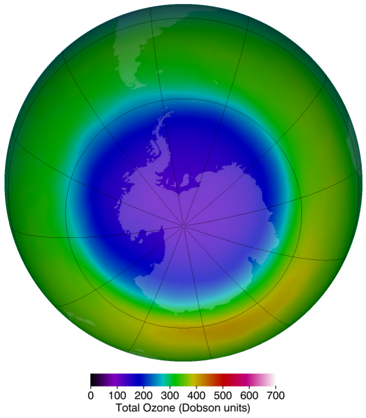
The Cl atom then catalyzes the breakdown of ozone as follows:

Equation:

$$\mathrm{Cl} + \mathrm{O_3} \rightarrow \mathrm{ClO} + \mathrm{O_2} \text{ and } \mathrm{ClO} + \mathrm{O_3} \rightarrow \mathrm{Cl} + 2\mathrm{O_2}.$$

A single chlorine atom could destroy ozone molecules for up to two years before being transported down to the surface. The CFCs are relatively stable and will contribute to ozone depletion for years to come. CFCs are found in refrigerants, air conditioning systems, foams, and aerosols.

International concern over this problem led to the establishment of the "Montreal Protocol" agreement (1987) to phase out CFC production in most countries. However, developing-country participation is needed if worldwide production and elimination of CFCs is to be achieved. Probably the largest contributor to CFC emissions today is India. But the protocol seems to be working, as there are signs of an ozone recovery. (See [link].)



This map of ozone concentration over Antarctica in October 2011 shows severe depletion suspected to be caused by CFCs. Less dramatic but more general depletion has been observed over northern latitudes, suggesting the effect is global. With less ozone, more ultraviolet radiation from the Sun reaches the surface, causing more damage. (credit: NASA Ozone Watch)

Benefits of UV Light

Besides the adverse effects of ultraviolet radiation, there are also benefits of exposure in nature and uses in technology. Vitamin D production in the skin (epidermis) results from exposure to UVB radiation, generally from sunlight. A number of studies indicate lack of vitamin D can result in the development of a range of cancers (prostate, breast, colon), so a certain amount of UV exposure is helpful. Lack of vitamin D is also linked to osteoporosis. Exposures (with no sunscreen) of 10 minutes a day to arms, face, and legs might be sufficient to provide the accepted dietary level. However, in the winter time north of about 37^{o} latitude, most UVB gets blocked by the atmosphere.

UV radiation is used in the treatment of infantile jaundice and in some skin conditions. It is also used in sterilizing workspaces and tools, and killing germs in a wide range of applications. It is

also used as an analytical tool to identify substances.

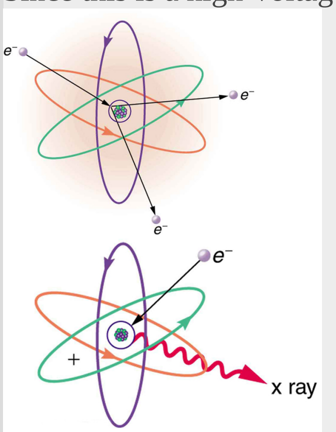
When exposed to ultraviolet, some substances, such as minerals, glow in characteristic visible wavelengths, a process called fluorescence. So-called black lights emit ultraviolet to cause posters and clothing to fluoresce in the visible. Ultraviolet is also used in special microscopes to detect details smaller than those observable with longer-wavelength visible-light microscopes.

Note:

Things Great and Small: A Submicroscopic View of X-Ray Production

X-rays can be created in a high-voltage discharge. They are emitted in the material struck by electrons in the discharge current. There are two mechanisms by which the electrons create X-rays.

The first method is illustrated in [link]. An electron is accelerated in an evacuated tube by a high positive voltage. The electron strikes a metal plate (e.g., copper) and produces X-rays. Since this is a high-voltage discharge, the electron gains sufficient energy to ionize the atom.

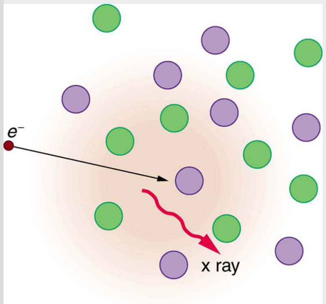


Artist’s conception of an electron ionizing an atom followed by the recapture of an electron and emission of an X-ray. An energetic electron strikes an atom and knocks an electron out of one of the orbits closest to the nucleus. Later, the atom captures another electron, and the energy released by

its fall into a low orbit generates a high-energy EM wave called an X-ray.

In the case shown, an inner-shell electron (one in an orbit relatively close to and tightly bound to the nucleus) is ejected. A short time later, another electron is captured and falls into the orbit in a single great plunge. The energy released by this fall is given to an EM wave known as an X-ray. Since the orbits of the atom are unique to the type of atom, the energy of the X-ray is characteristic of the atom, hence the name characteristic X-ray.

The second method by which an energetic electron creates an X-ray when it strikes a material is illustrated in [link]. The electron interacts with charges in the material as it penetrates. These collisions transfer kinetic energy from the electron to the electrons and atoms in the material.



Artist’s conception of an electron being slowed by collisions in a material and emitting X-ray radiation. This energetic electron makes numerous collisions with electrons and atoms in a material it penetrates. An accelerated charge radiates EM waves, a second method by which X-rays are created.

A loss of kinetic energy implies an acceleration, in this case decreasing the electron’s velocity. Whenever a charge is accelerated, it radiates EM waves. Given the high energy of the electron,

these EM waves can have high energy. We call them X-rays. Since the process is random, a broad spectrum of X-ray energy is emitted that is more characteristic of the electron energy than the type of material the electron encounters. Such EM radiation is called “bremsstrahlung” (German for “braking radiation”).

X-Rays

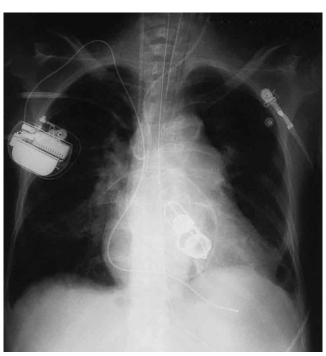
In the 1850s, scientists (such as Faraday) began experimenting with high-voltage electrical discharges in tubes filled with rarefied gases. It was later found that these discharges created an invisible, penetrating form of very high frequency electromagnetic radiation. This radiation was called an **X-ray**, because its identity and nature were unknown.

As described in Things Great and Small, there are two methods by which X-rays are created—both are submicroscopic processes and can be caused by high-voltage discharges. While the low-frequency end of the X-ray range overlaps with the ultraviolet, X-rays extend to much higher frequencies (and energies).

X-rays have adverse effects on living cells similar to those of ultraviolet radiation, and they have the additional liability of being more penetrating, affecting more than the surface layers of cells. Cancer and genetic defects can be induced by exposure to X-rays. Because of their effect on rapidly dividing cells, X-rays can also be used to treat and even cure cancer.

The widest use of X-rays is for imaging objects that are opaque to visible light, such as the human body or aircraft parts. In humans, the risk of cell damage is weighed carefully against the benefit of the diagnostic information obtained. However, questions have risen in recent years as to accidental overexposure of some people during CT scans—a mistake at least in part due to poor monitoring of radiation dose.

The ability of X-rays to penetrate matter depends on density, and so an X-ray image can reveal very detailed density information. [link] shows an example of the simplest type of X-ray image, an X-ray shadow on film. The amount of information in a simple X-ray image is impressive, but more sophisticated techniques, such as CT scans, can reveal three-dimensional information with details smaller than a millimeter.



This shadow X-ray image shows many interesting features, such as artificial heart valves, a pacemaker, and the wires used to close the sternum. (credit: P. P. Urone)

The use of X-ray technology in medicine is called radiology—an established and relatively cheap tool in comparison to more sophisticated technologies. Consequently, X-rays are widely available and used extensively in medical diagnostics. During World War I, mobile X-ray units, advocated by Madame Marie Curie, were used to diagnose soldiers.

Because they can have wavelengths less than 0.01 nm, X-rays can be scattered (a process called X-ray diffraction) to detect the shape of molecules and the structure of crystals. X-ray diffraction was crucial to Crick, Watson, and Wilkins in the determination of the shape of the double-helix DNA molecule.

X-rays are also used as a precise tool for trace-metal analysis in X-ray induced fluorescence, in which the energy of the X-ray emissions are related to the specific types of elements and amounts of materials present.

Gamma Rays

Soon after nuclear radioactivity was first detected in 1896, it was found that at least three distinct types of radiation were being emitted. The most penetrating nuclear radiation was called a **gamma ray (γ ray)** (again a name given because its identity and character were unknown), and it was later found to be an extremely high frequency electromagnetic wave.

In fact, γ rays are any electromagnetic radiation emitted by a nucleus. This can be from natural nuclear decay or induced nuclear processes in nuclear reactors and weapons. The lower end of the $\gamma\text{-ray}$ frequency range overlaps the upper end of the X-ray range, but γ rays can have the highest frequency of any electromagnetic radiation.

Gamma rays have characteristics identical to X-rays of the same frequency—they differ only in source. At higher frequencies, γ rays are more penetrating and more damaging to living tissue. They have many of the same uses as X-rays, including cancer therapy. Gamma radiation from radioactive materials is used in nuclear medicine.

[link] shows a medical image based on γ rays. Food spoilage can be greatly inhibited by exposing it to large doses of γ radiation, thereby obliterating responsible microorganisms. Damage to food cells through irradiation occurs as well, and the long-term hazards of

consuming radiation-preserved food are unknown and controversial for some groups. Both X-ray and γ-ray technologies are also used in scanning luggage at airports.

This is an image of the γ rays emitted by nuclei in a compound that is concentrated in the bones and eliminated through the kidneys. Bone cancer is evidenced by nonuniform concentration in similar

structures. For example, some ribs are darker than others. (credit: P. P. Urone)

Detecting Electromagnetic Waves from Space

A final note on star gazing. The entire electromagnetic spectrum is used by researchers for investigating stars, space, and time. As noted earlier, Penzias and Wilson detected microwaves to identify the background radiation originating from the Big Bang. Radio telescopes such as the Arecibo Radio Telescope in Puerto Rico and Parkes Observatory in Australia were designed to detect radio waves.

Infrared telescopes need to have their detectors cooled by liquid nitrogen to be able to gather useful signals. Since infrared radiation is predominantly from thermal agitation, if the detectors were not cooled, the vibrations of the molecules in the antenna would be stronger than the signal being collected.

The most famous of these infrared sensitive telescopes is the James Clerk Maxwell Telescope in Hawaii. The earliest telescopes, developed in the seventeenth century, were optical telescopes, collecting visible light. Telescopes in the ultraviolet, X-ray, and γ-ray regions are placed outside the atmosphere on satellites orbiting the Earth.

The Hubble Space Telescope (launched in 1990) gathers ultraviolet radiation as well as visible light. In the X-ray region, there is the Chandra X-ray Observatory (launched in 1999), and in the γ-ray region, there is the new Fermi Gamma-ray Space Telescope (launched in 2008—taking the place of the Compton Gamma Ray Observatory, 1991–2000.).

Note:
PhET Explorations: Color Vision
Make a whole rainbow by mixing red, green, and blue light. Change the wavelength of a monochromatic beam or filter white light. View the light as a solid beam, or see the individual photons.

Color Vision

Section Summary

- The relationship among the speed of propagation, wavelength, and frequency for any wave is given by $v_{\text{W}} = f\lambda$, so that for electromagnetic waves,
  Equation:
  $$c = f\lambda,$$
  where f is the frequency, λ is the wavelength, and c is the speed of light.
- The electromagnetic spectrum is separated into many categories and subcategories, based on the frequency and wavelength, source, and uses of the electromagnetic waves.
- Any electromagnetic wave produced by currents in wires is classified as a radio wave, the lowest frequency electromagnetic waves. Radio waves are divided into many types, depending on their applications, ranging up to microwaves at their highest frequencies.
- Infrared radiation lies below visible light in frequency and is produced by thermal motion and the vibration and rotation of atoms and molecules. Infrared's lower frequencies overlap with the highest-frequency microwaves.
- Visible light is largely produced by electronic transitions in atoms and molecules, and is defined as being detectable by the human eye. Its colors vary with frequency, from red at the lowest to violet at the highest.
- Ultraviolet radiation starts with frequencies just above violet in the visible range and is produced primarily by electronic transitions in atoms and molecules.
- X-rays are created in high-voltage discharges and by electron bombardment of metal targets. Their lowest frequencies overlap the ultraviolet range but extend to much higher values, overlapping at the high end with gamma rays.
- Gamma rays are nuclear in origin and are defined to include the highest-frequency electromagnetic radiation of any type.

Conceptual Questions

Exercise:

Problem:

If you live in a region that has a particular TV station, you can sometimes pick up some of its audio portion on your FM radio receiver. Explain how this is possible. Does it imply that TV audio is broadcast as FM?

Exercise:

Problem:

Explain why people who have the lens of their eye removed because of cataracts are able to see low-frequency ultraviolet.

Exercise:

Problem:

How do fluorescent soap residues make clothing look “brighter and whiter” in outdoor light? Would this be effective in candlelight?

Exercise:

Problem: Give an example of resonance in the reception of electromagnetic waves.

Exercise:

Problem:

Illustrate that the size of details of an object that can be detected with electromagnetic waves is related to their wavelength, by comparing details observable with two different types (for example, radar and visible light or infrared and X-rays).

Exercise:

Problem: Why don’t buildings block radio waves as completely as they do visible light?

Exercise:

Problem:

Make a list of some everyday objects and decide whether they are transparent or opaque to each of the types of electromagnetic waves.

Exercise:

Problem:

Your friend says that more patterns and colors can be seen on the wings of birds if viewed in ultraviolet light. Would you agree with your friend? Explain your answer.

Exercise:

Problem:

The rate at which information can be transmitted on an electromagnetic wave is proportional to the frequency of the wave. Is this consistent with the fact that laser telephone transmission at visible frequencies carries far more conversations per optical fiber than conventional electronic transmission in a wire? What is the implication for ELF radio communication with submarines?

Exercise:

Problem: Give an example of energy carried by an electromagnetic wave.

Exercise:

Problem:

In an MRI scan, a higher magnetic field requires higher frequency radio waves to resonate with the nuclear type whose density and location is being imaged. What effect does going to a larger magnetic field have on the most efficient antenna to broadcast those radio waves? Does it favor a smaller or larger antenna?

Exercise:

Problem:

Laser vision correction often uses an excimer laser that produces 193-nm electromagnetic radiation. This wavelength is extremely strongly absorbed by the cornea and ablates it in a manner that reshapes the cornea to correct vision defects. Explain how the strong absorption helps concentrate the energy in a thin layer and thus give greater accuracy in shaping the cornea. Also explain how this strong absorption limits damage to the lens and retina of the eye.

Problems & Exercises

Exercise:

Problem:

(a) Two microwave frequencies are authorized for use in microwave ovens: 900 and 2560 MHz. Calculate the wavelength of each. (b) Which frequency would produce smaller hot spots in foods due to interference effects?

Solution:

(a) 33.3 cm (900 MHz) 11.7 cm (2560 MHz)

(b) The microwave oven with the smaller wavelength would produce smaller hot spots in foods, corresponding to the one with the frequency 2560 MHz.

Exercise:

Problem:

(a) Calculate the range of wavelengths for AM radio given its frequency range is 540 to 1600 kHz. (b) Do the same for the FM frequency range of 88.0 to 108 MHz.

Exercise:

Problem:

A radio station utilizes frequencies between commercial AM and FM. What is the frequency of a 11.12-m-wavelength channel?

Solution:

26.96 MHz

Exercise:

Problem:

Find the frequency range of visible light, given that it encompasses wavelengths from 380 to 760 nm.

Exercise:

Problem:

Combing your hair leads to excess electrons on the comb. How fast would you have to move the comb up and down to produce red light?

Solution:

5.0×10^{14} Hz

Exercise:

Problem:

Electromagnetic radiation having a $15.0 - \mu\text{m}$ wavelength is classified as infrared radiation. What is its frequency?

Exercise:

Problem:

Approximately what is the smallest detail observable with a microscope that uses ultraviolet light of frequency 1.20×10^{15} Hz?

Solution:
Equation:

$$\lambda = \frac{c}{f} = \frac{3.00 \times 10^{8}\ \text{m/s}}{1.20 \times 10^{15}\ \text{Hz}} = 2.50 \times 10^{-7}\ \text{m}$$

Exercise:

Problem:

A radar used to detect the presence of aircraft receives a pulse that has reflected off an object 6×10^{-5} s after it was transmitted. What is the distance from the radar station to the reflecting object?

Exercise:

Problem:

Some radar systems detect the size and shape of objects such as aircraft and geological terrain. Approximately what is the smallest observable detail utilizing 500-MHz radar?

Solution:

0.600 m

Exercise:

Problem:

Determine the amount of time it takes for X-rays of frequency 3×10^{18} Hz to travel (a) 1 mm and (b) 1 cm.

Exercise:

Problem:

If you wish to detect details of the size of atoms (about 1×10^{-10} m) with electromagnetic radiation, it must have a wavelength of about this size. (a) What is its frequency? (b) What type of electromagnetic radiation might this be?

Solution:

(a) $f = \frac{c}{\lambda} = \frac{3.00\times10^{8}\text{ m/s}}{1\times10^{-10}\text{ m}} = 3\times10^{18}\text{ Hz}$

(b) X-rays

Exercise:

Problem:

If the Sun suddenly turned off, we would not know it until its light stopped coming. How long would that be, given that the Sun is 1.50×10^{11} m away?

Exercise:

Problem:

Distances in space are often quoted in units of light years, the distance light travels in one year. (a) How many meters is a light year? (b) How many meters is it to Andromeda, the nearest large galaxy, given that it is 2.00×10^{6} light years away? (c) The most distant galaxy yet discovered is 12.0×10^{9} light years away. How far is this in meters?

Exercise:

Problem:

A certain 50.0-Hz AC power line radiates an electromagnetic wave having a maximum electric field strength of 13.0 kV/m. (a) What is the wavelength of this very low frequency electromagnetic wave? (b) What is its maximum magnetic field strength?

Solution:

(a) $6.00 \times 10^6\ \text{m}$

(b) $4.33 \times 10^{-5}\ \text{T}$

Exercise:

Problem:

During normal beating, the heart creates a maximum 4.00-mV potential across 0.300 m of a person's chest, creating a 1.00-Hz electromagnetic wave. (a) What is the maximum electric field strength created? (b) What is the corresponding maximum magnetic field strength in the electromagnetic wave? (c) What is the wavelength of the electromagnetic wave?

Exercise:

Problem:

(a) The ideal size (most efficient) for a broadcast antenna with one end on the ground is one-fourth the wavelength ($\lambda/4$) of the electromagnetic radiation being sent out. If a new radio station has such an antenna that is 50.0 m high, what frequency does it broadcast most efficiently? Is this in the AM or FM band? (b) Discuss the analogy of the fundamental resonant mode of an air column closed at one end to the resonance of currents on an antenna that is one-fourth their wavelength.

Solution:

(a) 1.50×10^{6} Hz, AM band
(b) The resonance of currents on an antenna that is 1/4 their wavelength is analogous to the fundamental resonant mode of an air column closed at one end, since the tube also has a length equal to 1/4 the wavelength of the fundamental oscillation.

Exercise:

Problem:

(a) What is the wavelength of 100-MHz radio waves used in an MRI unit? (b) If the frequencies are swept over a ± 1.00 range centered on 100 MHz, what is the range of wavelengths broadcast?

Exercise:

Problem:

(a) What is the frequency of the 193-nm ultraviolet radiation used in laser eye surgery? (b) Assuming the accuracy with which this EM radiation can ablate the cornea is directly proportional to wavelength, how much more accurate can this UV be than the shortest visible wavelength of light?

Solution:

(a) $1.55 \times 10^{15}\ \text{Hz}$

(b) The shortest wavelength of visible light is 380 nm, so that
Equation:

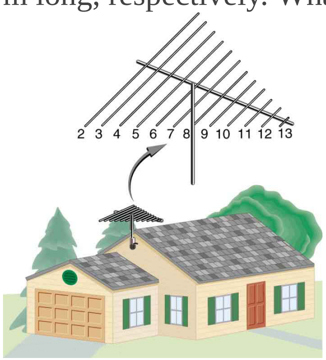
$$\frac{\lambda_{\text{visible}}}{\lambda_{\text{UV}}} = \frac{380\ \text{nm}}{193\ \text{nm}} = 1.97.$$

In other words, the UV radiation is 97% more accurate than the shortest wavelength of visible light, or almost twice as accurate!

Exercise:

Problem:

TV-reception antennas for VHF are constructed with cross wires supported at their centers, as shown in [link]. The ideal length for the cross wires is one-half the wavelength to be received, with the more expensive antennas having one for each channel. Suppose you measure the lengths of the wires for particular channels and find them to be 1.94 and 0.753 m long, respectively. What are the frequencies for these channels?



A television reception antenna has cross wires of various lengths to most efficiently

receive different
wavelengths.

Exercise:

Problem:

Conversations with astronauts on lunar walks had an echo that was used to estimate the distance to the Moon. The sound spoken by the person on Earth was transformed into a radio signal sent to the Moon, and transformed back into sound on a speaker inside the astronaut's space suit. This sound was picked up by the microphone in the space suit (intended for the astronaut's voice) and sent back to Earth as a radio echo of sorts. If the round-trip time was 2.60 s, what was the approximate distance to the Moon, neglecting any delays in the electronics?

Solution:

$3.90 \times 10^8 \text{ m}$

Exercise:

Problem:

Lunar astronauts placed a reflector on the Moon's surface, off which a laser beam is periodically reflected. The distance to the Moon is calculated from the round-trip time. (a) To what accuracy in meters can the distance to the Moon be determined, if this time can be measured to 0.100 ns? (b) What percent accuracy is this, given the average distance to the Moon is $3.84{\times}10^8 \text{ m}$?

Exercise:

Problem:

Radar is used to determine distances to various objects by measuring the round-trip time for an echo from the object. (a) How far away is the planet Venus if the echo time is 1000 s? (b) What is the echo time for a car 75.0 m from a Highway Police radar unit? (c) How accurately (in nanoseconds) must you be able to measure the echo time to an airplane 12.0 km away to determine its distance within 10.0 m?

Solution:

(a) $1.50 \times 10^{11} \text{ m}$

(b) $0.500\ \mu\text{s}$

(c) 66.7 ns

Exercise:

Problem: **Integrated Concepts**

(a) Calculate the ratio of the highest to lowest frequencies of electromagnetic waves the eye can see, given the wavelength range of visible light is from 380 to 760 nm. (b) Compare this with the ratio of highest to lowest frequencies the ear can hear.

Exercise:

Problem: **Integrated Concepts**

(a) Calculate the rate in watts at which heat transfer through radiation occurs (almost entirely in the infrared) from $1.0\ \mathrm{m}^2$ of the Earth's surface at night. Assume the emissivity is 0.90, the temperature of the Earth is $15^{\mathrm{o}}\mathrm{C}$, and that of outer space is 2.7 K. (b) Compare the intensity of this radiation with that coming to the Earth from the Sun during the day, which averages about $800\ \mathrm{W/m}^2$, only half of which is absorbed. (c) What is the maximum magnetic field strength in the outgoing radiation, assuming it is a continuous wave?

Solution:

(a) $-3.5\times10^2\ \mathrm{W/m}^2$

(b) 88%

(c) $1.7\ \mu\mathrm{T}$

Glossary

electromagnetic spectrum
: the full range of wavelengths or frequencies of electromagnetic radiation

radio waves
: electromagnetic waves with wavelengths in the range from 1 mm to 100 km; they are produced by currents in wires and circuits and by astronomical phenomena

microwaves
: electromagnetic waves with wavelengths in the range from 1 mm to 1 m; they can be produced by currents in macroscopic circuits and devices

thermal agitation
: the thermal motion of atoms and molecules in any object at a temperature above absolute zero, which causes them to emit and absorb radiation

radar
: a common application of microwaves. Radar can determine the distance to objects as diverse as clouds and aircraft, as well as determine the speed of a car or the intensity of a

rainstorm

infrared radiation (IR)
: a region of the electromagnetic spectrum with a frequency range that extends from just below the red region of the visible light spectrum up to the microwave region, or from $0.74\ \mu m$ to $300\ \mu m$

ultraviolet radiation (UV)
: electromagnetic radiation in the range extending upward in frequency from violet light and overlapping with the lowest X-ray frequencies, with wavelengths from 400 nm down to about 10 nm

visible light
: the narrow segment of the electromagnetic spectrum to which the normal human eye responds

amplitude modulation (AM)
: a method for placing information on electromagnetic waves by modulating the amplitude of a carrier wave with an audio signal, resulting in a wave with constant frequency but varying amplitude

extremely low frequency (ELF)
: electromagnetic radiation with wavelengths usually in the range of 0 to 300 Hz, but also about 1kHz

carrier wave
: an electromagnetic wave that carries a signal by modulation of its amplitude or frequency

frequency modulation (FM)
: a method of placing information on electromagnetic waves by modulating the frequency of a carrier wave with an audio signal, producing a wave of constant amplitude but varying frequency

TV
: video and audio signals broadcast on electromagnetic waves

very high frequency (VHF)
: TV channels utilizing frequencies in the two ranges of 54 to 88 MHz and 174 to 222 MHz

ultra-high frequency (UHF)
: TV channels in an even higher frequency range than VHF, of 470 to 1000 MHz

X-ray
: invisible, penetrating form of very high frequency electromagnetic radiation, overlapping both the ultraviolet range and the γ-ray range

gamma ray

(γ ray); extremely high frequency electromagnetic radiation emitted by the nucleus of an atom, either from natural nuclear decay or induced nuclear processes in nuclear reactors and weapons. The lower end of the γ-ray frequency range overlaps the upper end of the X-ray range, but γ rays can have the highest frequency of any electromagnetic radiation

Energy in Electromagnetic Waves

- Explain how the energy and amplitude of an electromagnetic wave are related.
- Given its power output and the heating area, calculate the intensity of a microwave oven's electromagnetic field, as well as its peak electric and magnetic field strengths

Anyone who has used a microwave oven knows there is energy in **electromagnetic waves**. Sometimes this energy is obvious, such as in the warmth of the summer sun. Other times it is subtle, such as the unfelt energy of gamma rays, which can destroy living cells.

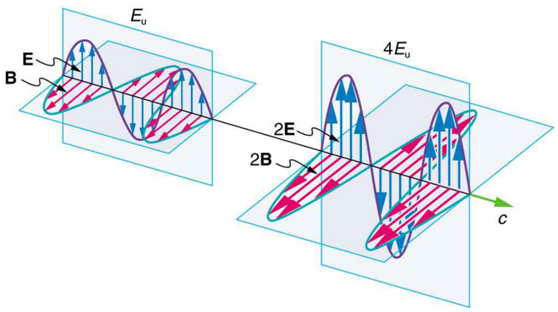
Electromagnetic waves can bring energy into a system by virtue of their **electric and magnetic fields**. These fields can exert forces and move charges in the system and, thus, do work on them. If the frequency of the electromagnetic wave is the same as the natural frequencies of the system (such as microwaves at the resonant frequency of water molecules), the transfer of energy is much more efficient.

Note:

Connections: Waves and Particles

The behavior of electromagnetic radiation clearly exhibits wave characteristics. But we shall find in later modules that at high frequencies, electromagnetic radiation also exhibits particle characteristics. These particle characteristics will be used to explain more of the properties of the electromagnetic spectrum and to introduce the formal study of modern physics.

Another startling discovery of modern physics is that particles, such as electrons and protons, exhibit wave characteristics. This simultaneous sharing of wave and particle properties for all submicroscopic entities is one of the great symmetries in nature.



Energy carried by a wave is proportional to its amplitude squared. With electromagnetic waves, larger E-fields and B-fields exert larger forces and can do more work.

But there is energy in an electromagnetic wave, whether it is absorbed or not. Once created, the fields carry energy away from a source. If absorbed, the field strengths are diminished and anything left travels on. Clearly, the larger the strength of the electric and magnetic fields, the more work they can do and the greater the energy the electromagnetic wave carries.

A wave's energy is proportional to its **amplitude** squared (E^2 or B^2). This is true for waves on guitar strings, for water waves, and for sound waves, where amplitude is proportional to pressure. In electromagnetic waves, the amplitude is the **maximum field strength** of the electric and magnetic fields. (See [link].)

Thus the energy carried and the **intensity** I of an electromagnetic wave is proportional to E^2 and B^2. In fact, for a continuous sinusoidal electromagnetic wave, the average intensity I_{ave} is given by
Equation:

$$I_{\text{ave}} = \frac{c\varepsilon_0 E_0^2}{2},$$

where c is the speed of light, ε_0 is the permittivity of free space, and E_0 is the maximum electric field strength; intensity, as always, is power per unit area (here in $\mathrm{W/m^2}$).

The average intensity of an electromagnetic wave I_{ave} can also be expressed in terms of the magnetic field strength by using the relationship $B = E/c$, and the fact that $\varepsilon_0 = 1/\mu_0 c^2$, where μ_0 is the permeability of free space. Algebraic manipulation produces the relationship
Equation:

$$I_{\text{ave}} = \frac{\mathrm{cB}_0^2}{2\mu_0},$$

where B_0 is the maximum magnetic field strength.

One more expression for I_{ave} in terms of both electric and magnetic field strengths is useful. Substituting the fact that $c \cdot B_0 = E_0$, the previous expression becomes
Equation:

$$I_{\text{ave}} = \frac{E_0 B_0}{2\mu_0}.$$

Whichever of the three preceding equations is most convenient can be used, since they are really just different versions of the same principle: Energy in a wave is related to amplitude squared. Furthermore, since these equations are based on the assumption that the electromagnetic waves are sinusoidal, peak intensity is twice the average; that is, $I_0 = 2I_{\text{ave}}$.

Example:
Calculate Microwave Intensities and Fields
On its highest power setting, a certain microwave oven projects 1.00 kW of microwaves onto a 30.0 by 40.0 cm area. (a) What is the intensity in

W/m^2? (b) Calculate the peak electric field strength E_0 in these waves. (c) What is the peak magnetic field strength B_0?

Strategy

In part (a), we can find intensity from its definition as power per unit area. Once the intensity is known, we can use the equations below to find the field strengths asked for in parts (b) and (c).

Solution for (a)

Entering the given power into the definition of intensity, and noting the area is 0.300 by 0.400 m, yields

Equation:

$$I = \frac{P}{A} = \frac{1.00\ \text{kW}}{0.300\ \text{m}\ \times\ 0.400\ \text{m}}.$$

Here $I = I_{\text{ave}}$, so that

Equation:

$$I_{\text{ave}} = \frac{1000\ \text{W}}{0.120\ \text{m}^2} = 8.33 \times 10^3\ \text{W/m}^2.$$

Note that the peak intensity is twice the average:

Equation:

$$I_0 = 2I_{\text{ave}} = 1.67 \times 10^4\ \text{W/m}^2.$$

Solution for (b)

To find E_0, we can rearrange the first equation given above for I_{ave} to give

Equation:

$$E_0 = \left(\frac{2I_{\text{ave}}}{c\varepsilon_0}\right)^{1/2}.$$

Entering known values gives

Equation:

$$\begin{aligned} E_0 &= \sqrt{\frac{2(8.33\times10^3\ \text{W/m}^2)}{(3.00\times10^8\ \text{m/s})(8.85\times10^{-12}\ \text{C}^2/\text{N}\cdot\text{m}^2)}} \\ &= 2.51\times10^3\ \text{V/m}. \end{aligned}$$

Solution for (c)

Perhaps the easiest way to find magnetic field strength, now that the electric field strength is known, is to use the relationship given by

Equation:

$$B_0 = \frac{E_0}{c}.$$

Entering known values gives

Equation:

$$\begin{aligned} B_0 &= \frac{2.51\times10^3\ \text{V/m}}{3.0\times10^8\ \text{m/s}} \\ &= 8.35\times10^{-6}\ \text{T}. \end{aligned}$$

Discussion

As before, a relatively strong electric field is accompanied by a relatively weak magnetic field in an electromagnetic wave, since $B = E/c$, and c is a large number.

Section Summary

- The energy carried by any wave is proportional to its amplitude squared. For electromagnetic waves, this means intensity can be expressed as

  Equation:

  $$I_{\text{ave}} = \frac{c\varepsilon_0 E_0^2}{2},$$

where I_{ave} is the average intensity in $\mathrm{W/m^2}$, and E_0 is the maximum electric field strength of a continuous sinusoidal wave.

- This can also be expressed in terms of the maximum magnetic field strength B_0 as
Equation:

$$I_{\mathrm{ave}} = \frac{\mathrm{cB_0^2}}{2\mathrm{\mu_0}}$$

and in terms of both electric and magnetic fields as
Equation:

$$I_{\mathrm{ave}} = \frac{E_0 B_0}{2\mathrm{\mu_0}}.$$

- The three expressions for I_{ave} are all equivalent.

Problems & Exercises

Exercise:

Problem:

What is the intensity of an electromagnetic wave with a peak electric field strength of 125 V/m?

Solution:
Equation:

$$\begin{aligned} I &= \frac{c\varepsilon_0 E_0^2}{2} \\ &= \frac{\left(3.00\times10^8\ \mathrm{m/s}\right)\left(8.85\times10^{-12}\mathrm{C^2/N\cdot m^2}\right)\left(125\ \mathrm{V/m}\right)^2}{2} \\ &= 20.7\ \mathrm{W/m^2} \end{aligned}$$

Exercise:

Problem:

Find the intensity of an electromagnetic wave having a peak magnetic field strength of 4.00×10^{-9} T.

Exercise:

Problem:

Assume the helium-neon lasers commonly used in student physics laboratories have power outputs of 0.250 mW. (a) If such a laser beam is projected onto a circular spot 1.00 mm in diameter, what is its intensity? (b) Find the peak magnetic field strength. (c) Find the peak electric field strength.

Solution:

(a) $I = \frac{P}{A} = \frac{P}{\pi r^2} = \frac{0.250\times10^{-3}\text{ W}}{\pi\left(0.500\times10^{-3}\text{ m}\right)^2} = 318\text{ W/m}^2$

(b)
$$\begin{aligned} I_{\text{ave}} &= \frac{\text{cB}_0^2}{2\mu_0} \Rightarrow B_0 = \left(\frac{2\mu_0 I}{c}\right)^{1/2} \\ &= \left(\frac{2\left(4\pi\times10^{-7}\text{ T}\cdot\text{m/A}\right)\left(318.3\text{ W/m}^2\right)}{3.00\times10^{8}\text{ m/s}}\right)^{1/2} \\ &= 1.63\times10^{-6}\text{ T} \end{aligned}$$

(c)
$$\begin{aligned} E_0 &= \text{cB}_0 = \left(3.00\times10^{8}\text{ m/s}\right)\left(1.633\times10^{-6}\text{ T}\right) \\ &= 4.90\times10^{2}\text{ V/m} \end{aligned}$$

Exercise:

Problem:

An AM radio transmitter broadcasts 50.0 kW of power uniformly in all directions. (a) Assuming all of the radio waves that strike the ground are completely absorbed, and that there is no absorption by the atmosphere or other objects, what is the intensity 30.0 km away? (Hint: Half the power will be spread over the area of a hemisphere.) (b) What is the maximum electric field strength at this distance?

Exercise:

Problem:

Suppose the maximum safe intensity of microwaves for human exposure is taken to be $1.00\ \text{W/m}^2$. (a) If a radar unit leaks 10.0 W of microwaves (other than those sent by its antenna) uniformly in all directions, how far away must you be to be exposed to an intensity considered to be safe? Assume that the power spreads uniformly over the area of a sphere with no complications from absorption or reflection. (b) What is the maximum electric field strength at the safe intensity? (Note that early radar units leaked more than modern ones do. This caused identifiable health problems, such as cataracts, for people who worked near them.)

Solution:

(a) 89.2 cm

(b) 27.4 V/m

Exercise:

Problem:

A 2.50-m-diameter university communications satellite dish receives TV signals that have a maximum electric field strength (for one channel) of $7.50\ \mu\text{V/m}$. (See [link].) (a) What is the intensity of this wave? (b) What is the power received by the antenna? (c) If the orbiting satellite broadcasts uniformly over an area of $1.50 \times 10^{13}\ \text{m}^2$ (a large fraction of North America), how much power does it radiate?

Satellite dishes receive TV signals sent from orbit. Although the signals are quite weak, the receiver can detect them by being tuned to resonate at their frequency.

Exercise:

Problem:

Lasers can be constructed that produce an extremely high intensity electromagnetic wave for a brief time—called pulsed lasers. They are used to ignite nuclear fusion, for example. Such a laser may produce an electromagnetic wave with a maximum electric field strength of $1.00 \times 10^{11}\ \text{V/m}$ for a time of 1.00 ns. (a) What is the maximum magnetic field strength in the wave? (b) What is the intensity of the beam? (c) What energy does it deliver on a 1.00-mm^2 area?

Solution:

(a) 333 T

(b) $1.33 \times 10^{19}\ \text{W/m}^2$

(c) 13.3 kJ

Exercise:

Problem:

Show that for a continuous sinusoidal electromagnetic wave, the peak intensity is twice the average intensity ($I_0 = 2I_{\text{ave}}$), using either the fact that $E_0 = \sqrt{2}E_{\text{rms}}$, or $B_0 = \sqrt{2}B_{\text{rms}}$, where rms means average (actually root mean square, a type of average).

Exercise:

Problem:

Suppose a source of electromagnetic waves radiates uniformly in all directions in empty space where there are no absorption or interference effects. (a) Show that the intensity is inversely proportional to r^2, the distance from the source squared. (b) Show that the magnitudes of the electric and magnetic fields are inversely proportional to r.

Solution:

(a) $I = \frac{P}{A} = \frac{P}{4\pi r^2} \propto \frac{1}{r^2}$

(b) $I \propto E_0^2, B_0^2 \Rightarrow E_0^2, B_0^2 \propto \frac{1}{r^2} \Rightarrow E_0, B_0 \propto \frac{1}{r}$

Exercise:

Problem: Integrated Concepts

An LC circuit with a 5.00-pF capacitor oscillates in such a manner as to radiate at a wavelength of 3.30 m. (a) What is the resonant frequency? (b) What inductance is in series with the capacitor?

Exercise:

Problem: Integrated Concepts

What capacitance is needed in series with an $800 - \mu\mathrm{H}$ inductor to form a circuit that radiates a wavelength of 196 m?

Solution:

13.5 pF

Exercise:

Problem: Integrated Concepts

Police radar determines the speed of motor vehicles using the same Doppler-shift technique employed for ultrasound in medical diagnostics. Beats are produced by mixing the double Doppler-shifted echo with the original frequency. If 1.50×10^9-Hz microwaves are used and a beat frequency of 150 Hz is produced, what is the speed of the vehicle? (Assume the same Doppler-shift formulas are valid with the speed of sound replaced by the speed of light.)

Exercise:

Problem: Integrated Concepts

Assume the mostly infrared radiation from a heat lamp acts like a continuous wave with wavelength $1.50\ \mu\text{m}$. (a) If the lamp's 200-W output is focused on a person's shoulder, over a circular area 25.0 cm in diameter, what is the intensity in W/m^2? (b) What is the peak electric field strength? (c) Find the peak magnetic field strength. (d) How long will it take to increase the temperature of the 4.00-kg shoulder by $2.00^{\circ}\ \text{C}$, assuming no other heat transfer and given that its specific heat is $3.47 \times 10^3\ \text{J/kg}\cdot{}^{\circ}\text{C}$?

Solution:

(a) $4.07\ \text{kW/m}^2$

(b) 1.75 kV/m

(c) $5.84\ \mu\text{T}$

(d) 2 min 19 s

Exercise:

Problem: Integrated Concepts

On its highest power setting, a microwave oven increases the temperature of 0.400 kg of spaghetti by 45.0°C in 120 s. (a) What was the rate of power absorption by the spaghetti, given that its specific heat is $3.76 \times 10^3\ \text{J/kg}\cdot{}^{\circ}\text{C}$? (b) Find the average intensity of the microwaves, given that they are absorbed over a circular area 20.0 cm in diameter. (c) What is the peak electric field strength of the microwave? (d) What is its peak magnetic field strength?

Exercise:

Problem: Integrated Concepts

Electromagnetic radiation from a 5.00-mW laser is concentrated on a 1.00-mm^2 area. (a) What is the intensity in W/m^2? (b) Suppose a 2.00-nC static charge is in the beam. What is the maximum electric

force it experiences? (c) If the static charge moves at 400 m/s, what maximum magnetic force can it feel?

Solution:

(a) $5.00\times10^{3}\ \mathrm{W/m^{2}}$

(b) $3.88\times10^{-6}\ \mathrm{N}$

(c) $5.18\times10^{-12}\ \mathrm{N}$

Exercise:

Problem: Integrated Concepts

A 200-turn flat coil of wire 30.0 cm in diameter acts as an antenna for FM radio at a frequency of 100 MHz. The magnetic field of the incoming electromagnetic wave is perpendicular to the coil and has a maximum strength of $1.00\times10^{-12}\ \mathrm{T}$. (a) What power is incident on the coil? (b) What average emf is induced in the coil over one-fourth of a cycle? (c) If the radio receiver has an inductance of $2.50\ \mu\mathrm{H}$, what capacitance must it have to resonate at 100 MHz?

Exercise:

Problem: Integrated Concepts

If electric and magnetic field strengths vary sinusoidally in time, being zero at $t=0$, then $E=E_0\ \sin\ 2\pi\mathrm{ft}$ and $B=B_0\ \sin\ 2\pi\mathrm{ft}$. Let $f=1.00\ \mathrm{GHz}$ here. (a) When are the field strengths first zero? (b) When do they reach their most negative value? (c) How much time is needed for them to complete one cycle?

Solution:

(a) $t=0$

(b) $7.50\times10^{-10}\ \mathrm{s}$

(c) 1.00×10^{-9} s

Exercise:

Problem: **Unreasonable Results**

A researcher measures the wavelength of a 1.20-GHz electromagnetic wave to be 0.500 m. (a) Calculate the speed at which this wave propagates. (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Exercise:

Problem: **Unreasonable Results**

The peak magnetic field strength in a residential microwave oven is 9.20×10^{-5} T. (a) What is the intensity of the microwave? (b) What is unreasonable about this result? (c) What is wrong about the premise?

Solution:

(a) $1.01 \times 10^{6}\ \mathrm{W/m^2}$

(b) Much too great for an oven.

(c) The assumed magnetic field is unreasonably large.

Exercise:

Problem: **Unreasonable Results**

An LC circuit containing a 2.00-H inductor oscillates at such a frequency that it radiates at a 1.00-m wavelength. (a) What is the capacitance of the circuit? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Exercise:

Problem: **Unreasonable Results**

An LC circuit containing a 1.00-pF capacitor oscillates at such a frequency that it radiates at a 300-nm wavelength. (a) What is the inductance of the circuit? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) $2.53 \times 10^{-20}\ \mathrm{H}$

(b) L is much too small.

(c) The wavelength is unreasonably small.

Exercise:

Problem: Create Your Own Problem

Consider electromagnetic fields produced by high voltage power lines. Construct a problem in which you calculate the intensity of this electromagnetic radiation in $\mathrm{W/m^2}$ based on the measured magnetic field strength of the radiation in a home near the power lines. Assume these magnetic field strengths are known to average less than a $\mu\mathrm{T}$. The intensity is small enough that it is difficult to imagine mechanisms for biological damage due to it. Discuss how much energy may be radiating from a section of power line several hundred meters long and compare this to the power likely to be carried by the lines. An idea of how much power this is can be obtained by calculating the approximate current responsible for $\mu\mathrm{T}$ fields at distances of tens of meters.

Exercise:

Problem: Create Your Own Problem

Consider the most recent generation of residential satellite dishes that are a little less than half a meter in diameter. Construct a problem in which you calculate the power received by the dish and the maximum electric field strength of the microwave signals for a single channel

received by the dish. Among the things to be considered are the power broadcast by the satellite and the area over which the power is spread, as well as the area of the receiving dish.

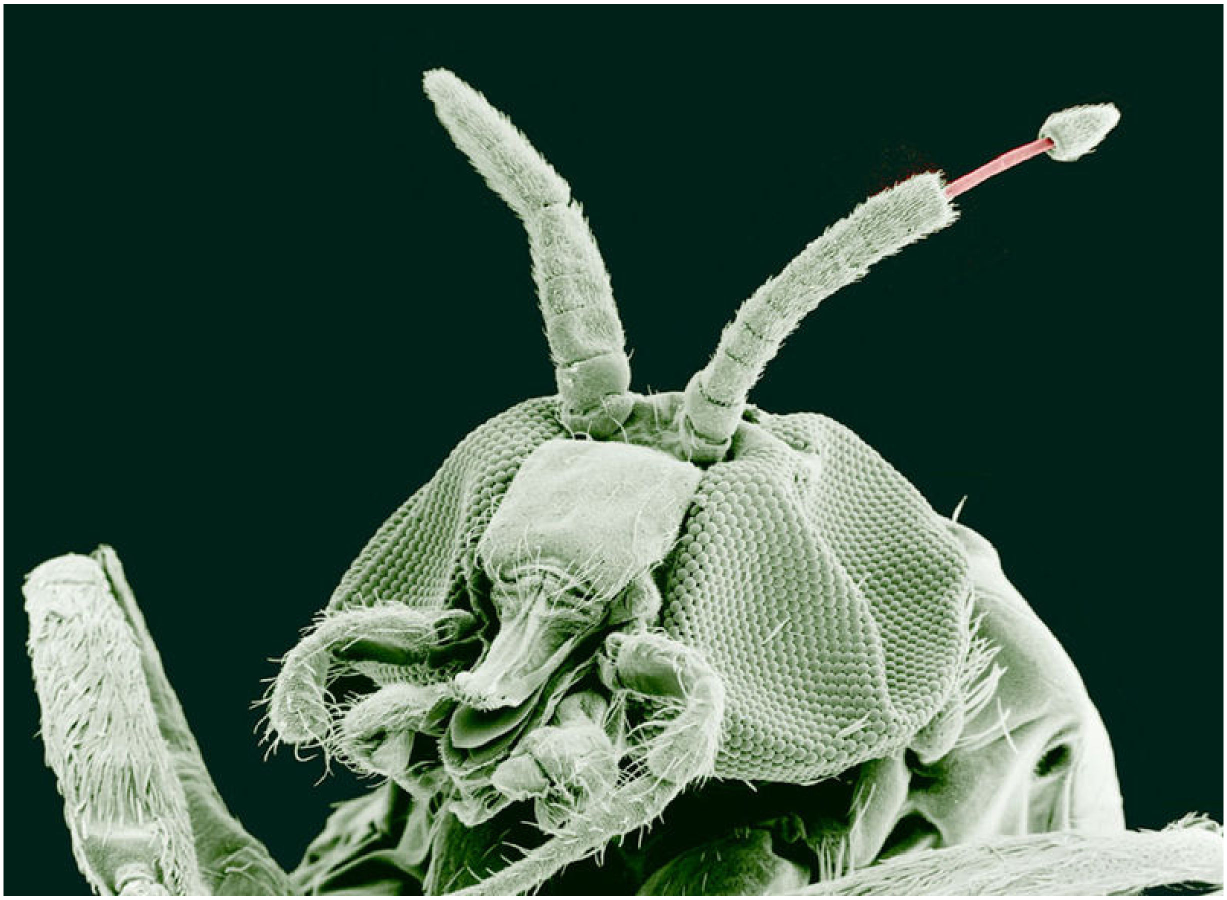
Glossary

maximum field strength
: the maximum amplitude an electromagnetic wave can reach, representing the maximum amount of electric force and/or magnetic flux that the wave can exert

intensity
: the power of an electric or magnetic field per unit area, for example, Watts per square meter

Introduction to Quantum Physics
class="introduction"

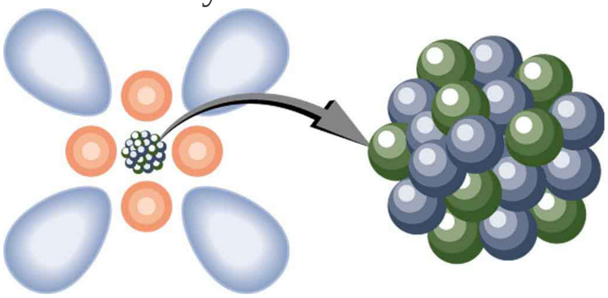
A black fly imaged by an electron microscope is as monstrous as any science-fiction creature. (credit: U.S. Department of Agriculture via Wikimedia Commons)



Quantum mechanics is the branch of physics needed to deal with submicroscopic objects. Because these objects are smaller than we can observe directly with our senses and generally must be observed with the aid of instruments, parts of quantum mechanics seem as foreign and bizarre as parts of relativity. But, like relativity, quantum mechanics has been shown to be valid—truth is often stranger than fiction.

Certain aspects of quantum mechanics are familiar to us. We accept as fact that matter is composed of atoms, the smallest unit of an element, and that these atoms combine to form molecules, the smallest unit of a compound. (See [link].) While we cannot see the individual water molecules in a stream, for example, we are aware that this is because molecules are so small and so numerous in that stream. When introducing atoms, we commonly say that electrons orbit atoms in discrete shells around a tiny nucleus, itself composed of smaller particles called protons and neutrons. We are also aware that electric charge comes in tiny units carried almost entirely by electrons and protons. As with water molecules in a stream, we

do not notice individual charges in the current through a lightbulb, because the charges are so small and so numerous in the macroscopic situations we sense directly.



Atoms and their substructure are familiar examples of objects that require quantum mechanics to be fully explained. Certain of their characteristics, such as the discrete electron shells, are classical physics explanations. In quantum mechanics we conceptualize discrete "electron clouds" around the nucleus.

Note:
Making Connections: Realms of Physics
Classical physics is a good approximation of modern physics under conditions first discussed in the The Nature of Science and Physics. Quantum mechanics is valid in general, and it must be used rather than classical physics to describe small objects, such as atoms.

Atoms, molecules, and fundamental electron and proton charges are all examples of physical entities that are **quantized**—that is, they appear only in certain discrete values and do not have every conceivable value.

Quantized is the opposite of continuous. We cannot have a fraction of an atom, or part of an electron’s charge, or 14-1/3 cents, for example. Rather, everything is built of integral multiples of these substructures. Quantum physics is the branch of physics that deals with small objects and the quantization of various entities, including energy and angular momentum. Just as with classical physics, quantum physics has several subfields, such as mechanics and the study of electromagnetic forces. The **correspondence principle** states that in the classical limit (large, slow-moving objects), **quantum mechanics** becomes the same as classical physics. In this chapter, we begin the development of quantum mechanics and its description of the strange submicroscopic world. In later chapters, we will examine many areas, such as atomic and nuclear physics, in which quantum mechanics is crucial.

Glossary

quantized
: the fact that certain physical entities exist only with particular discrete values and not every conceivable value

correspondence principle
: in the classical limit (large, slow-moving objects), quantum mechanics becomes the same as classical physics

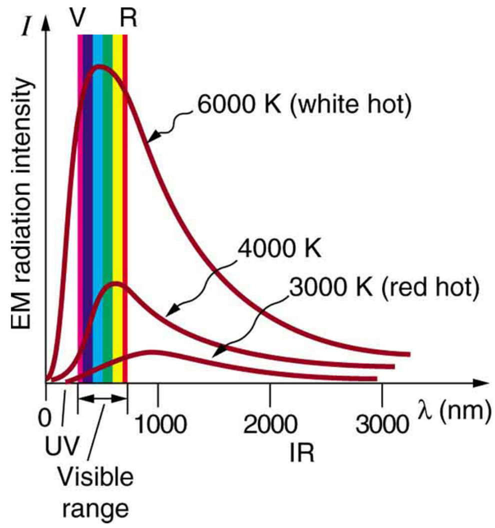
quantum mechanics
: the branch of physics that deals with small objects and with the quantization of various entities, especially energy

Quantization of Energy

- Explain Max Planck’s contribution to the development of quantum mechanics.
- Explain why atomic spectra indicate quantization.

Planck’s Contribution

Energy is quantized in some systems, meaning that the system can have only certain energies and not a continuum of energies, unlike the classical case. This would be like having only certain speeds at which a car can travel because its kinetic energy can have only certain values. We also find that some forms of energy transfer take place with discrete lumps of energy. While most of us are familiar with the quantization of matter into lumps called atoms, molecules, and the like, we are less aware that energy, too, can be quantized. Some of the earliest clues about the necessity of quantum mechanics over classical physics came from the quantization of energy.



Graphs of blackbody radiation (from an ideal radiator) at three different radiator temperatures. The intensity or rate of

radiation emission increases dramatically with temperature, and the peak of the spectrum shifts toward the visible and ultraviolet parts of the spectrum. The shape of the spectrum cannot be described with classical physics.

Where is the quantization of energy observed? Let us begin by considering the emission and absorption of electromagnetic (EM) radiation. The EM spectrum radiated by a hot solid is linked directly to the solid's temperature. (See [link].) An ideal radiator is one that has an emissivity of 1 at all wavelengths and, thus, is jet black. Ideal radiators are therefore called **blackbodies**, and their EM radiation is called **blackbody radiation**. It was discussed that the total intensity of the radiation varies as T^4, the fourth power of the absolute temperature of the body, and that the peak of the spectrum shifts to shorter wavelengths at higher temperatures. All of this seems quite continuous, but it was the curve of the spectrum of intensity versus wavelength that gave a clue that the energies of the atoms in the solid are quantized. In fact, providing a theoretical explanation for the experimentally measured shape of the spectrum was a mystery at the turn of the century. When this "ultraviolet catastrophe" was eventually solved, the answers led to new technologies such as computers and the sophisticated imaging techniques described in earlier chapters. Once again, physics as an enabling science changed the way we live.

The German physicist Max Planck (1858–1947) used the idea that atoms and molecules in a body act like oscillators to absorb and emit radiation. The energies of the oscillating atoms and molecules had to be quantized to correctly describe the shape of the blackbody spectrum. Planck deduced that the energy of an oscillator having a frequency f is given by

Equation:

$$E = \left(n + \frac{1}{2}\right)\text{hf}.$$

Here n is any nonnegative integer (0, 1, 2, 3, …). The symbol h stands for **Planck's constant**, given by
Equation:

$$h = 6.626 \times 10^{-34}\ \text{J} \cdot \text{s}.$$

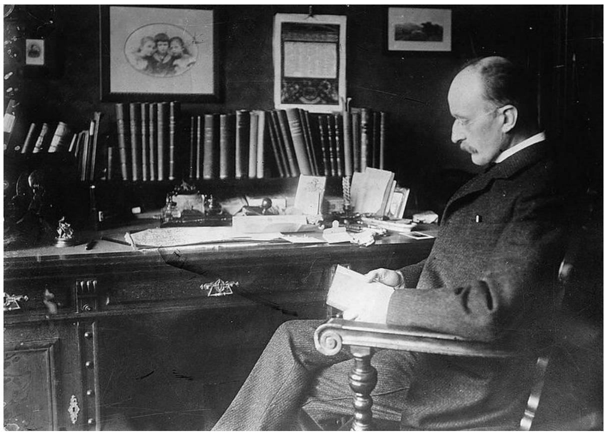
The equation $E = \left(n + \frac{1}{2}\right)\text{hf}$ means that an oscillator having a frequency f (emitting and absorbing EM radiation of frequency f) can have its energy increase or decrease only in *discrete* steps of size
Equation:

$$\Delta E = \text{hf}.$$

It might be helpful to mention some macroscopic analogies of this quantization of energy phenomena. This is like a pendulum that has a characteristic oscillation frequency but can swing with only certain amplitudes. Quantization of energy also resembles a standing wave on a string that allows only particular harmonics described by integers. It is also similar to going up and down a hill using discrete stair steps rather than being able to move up and down a continuous slope. Your potential energy takes on discrete values as you move from step to step.

Using the quantization of oscillators, Planck was able to correctly describe the experimentally known shape of the blackbody spectrum. This was the first indication that energy is sometimes quantized on a small scale and earned him the Nobel Prize in Physics in 1918. Although Planck's theory comes from observations of a macroscopic object, its analysis is based on atoms and molecules. It was such a revolutionary departure from classical physics that Planck himself was reluctant to accept his own idea that energy states are not continuous. The general acceptance of Planck's energy quantization was greatly enhanced by Einstein's explanation of the photoelectric effect (discussed in the next section), which took energy

quantization a step further. Planck was fully involved in the development of both early quantum mechanics and relativity. He quickly embraced Einstein's special relativity, published in 1905, and in 1906 Planck was the first to suggest the correct formula for relativistic momentum, $p = \gamma mu$.



The German physicist Max Planck had a major influence on the early development of quantum mechanics, being the first to recognize that energy is sometimes quantized. Planck also made important contributions to special relativity and classical physics. (credit: Library of Congress, Prints and Photographs Division via Wikimedia Commons)

Note that Planck's constant h is a very small number. So for an infrared frequency of 10^{14} Hz being emitted by a blackbody, for example, the difference between energy levels is only $\Delta E = \text{hf}{=}(6.63 \times 10^{-34}\ \text{J·s})(10^{14}\ \text{Hz}){=}\ 6.63 \times 10^{-20}\ \text{J}$, or about 0.4 eV. This 0.4 eV of energy is significant compared with typical atomic

energies, which are on the order of an electron volt, or thermal energies, which are typically fractions of an electron volt. But on a macroscopic or classical scale, energies are typically on the order of joules. Even if macroscopic energies are quantized, the quantum steps are too small to be noticed. This is an example of the correspondence principle. For a large object, quantum mechanics produces results indistinguishable from those of classical physics.

Atomic Spectra

Now let us turn our attention to the *emission and absorption of EM radiation by gases*. The Sun is the most common example of a body containing gases emitting an EM spectrum that includes visible light. We also see examples in neon signs and candle flames. Studies of emissions of hot gases began more than two centuries ago, and it was soon recognized that these emission spectra contained huge amounts of information. The type of gas and its temperature, for example, could be determined. We now know that these EM emissions come from electrons transitioning between energy levels in individual atoms and molecules; thus, they are called **atomic spectra**. Atomic spectra remain an important analytical tool today. [link] shows an example of an emission spectrum obtained by passing an electric discharge through a material. One of the most important characteristics of these spectra is that they are discrete. By this we mean that only certain wavelengths, and hence frequencies, are emitted. This is called a line spectrum. If frequency and energy are associated as $\Delta E = \mathrm{hf}$, the energies of the electrons in the emitting atoms and molecules are quantized. This is discussed in more detail later in this chapter.

Emission spectrum of oxygen. When an electrical discharge is passed through a substance, its atoms and molecules absorb energy, which is reemitted as EM radiation. The discrete nature of these emissions implies that the energy states of the atoms

and molecules are quantized. Such atomic spectra were used as analytical tools for many decades before it was understood why they are quantized. (credit: Teravolt, Wikimedia Commons)

It was a major puzzle that atomic spectra are quantized. Some of the best minds of 19th-century science failed to explain why this might be. Not until the second decade of the 20th century did an answer based on quantum mechanics begin to emerge. Again a macroscopic or classical body of gas was involved in the studies, but the effect, as we shall see, is due to individual atoms and molecules.

Note:
PhET Explorations: Models of the Hydrogen Atom
How did scientists figure out the structure of atoms without looking at them? Try out different models by shooting light at the atom. Check how the prediction of the model matches the experimental results.

https://archive.cnx.org/specials/d77cc1d0-33e4-11e6-b016-6726afecd2be/hydrogen-atom/#sim-hydrogen-atom

Section Summary

- The first indication that energy is sometimes quantized came from blackbody radiation, which is the emission of EM radiation by an object with an emissivity of 1.
- Planck recognized that the energy levels of the emitting atoms and molecules were quantized, with only the allowed values of $E = \left(n + \frac{1}{2}\right)\text{hf}$, where n is any non-negative integer (0, 1, 2, 3, …).
- h is Planck’s constant, whose value is $h = 6.626 \times 10^{–34}\ \text{J} \cdot \text{s}$.
- Thus, the oscillatory absorption and emission energies of atoms and molecules in a blackbody could increase or decrease only in steps of

size $\Delta E = \text{hf}$ where f is the frequency of the oscillatory nature of the absorption and emission of EM radiation.

- Another indication of energy levels being quantized in atoms and molecules comes from the lines in atomic spectra, which are the EM emissions of individual atoms and molecules.

Conceptual Questions

Exercise:

Problem:

Give an example of a physical entity that is quantized. State specifically what the entity is and what the limits are on its values.

Exercise:

Problem:

Give an example of a physical entity that is not quantized, in that it is continuous and may have a continuous range of values.

Exercise:

Problem:

What aspect of the blackbody spectrum forced Planck to propose quantization of energy levels in its atoms and molecules?

Exercise:

Problem:

If Planck's constant were large, say 10^{34} times greater than it is, we would observe macroscopic entities to be quantized. Describe the motions of a child's swing under such circumstances.

Exercise:

Problem: Why don't we notice quantization in everyday events?

Problems & Exercises

Exercise:

Problem:

A LiBr molecule oscillates with a frequency of $1.7 \times 10^{13}\ \mathrm{Hz}$. (a) What is the difference in energy in eV between allowed oscillator states? (b) What is the approximate value of n for a state having an energy of 1.0 eV?

Solution:

(a) 0.070 eV

(b) 14

Exercise:

Problem:

The difference in energy between allowed oscillator states in HBr molecules is 0.330 eV. What is the oscillation frequency of this molecule?

Exercise:

Problem:

A physicist is watching a 15-kg orangutan at a zoo swing lazily in a tire at the end of a rope. He (the physicist) notices that each oscillation takes 3.00 s and hypothesizes that the energy is quantized. (a) What is the difference in energy in joules between allowed oscillator states? (b) What is the value of n for a state where the energy is 5.00 J? (c) Can the quantization be observed?

Solution:

(a) 2.21×10^{34} J

(b) 2.26×10^{34}

(c) No

Glossary

blackbody
: an ideal radiator, which can radiate equally well at all wavelengths

blackbody radiation
: the electromagnetic radiation from a blackbody

Planck's constant
: $h = 6.626 \times 10^{-34}\ \mathrm{J} \cdot \mathrm{s}$

atomic spectra
: the electromagnetic emission from atoms and molecules

The Photoelectric Effect

- Describe a typical photoelectric-effect experiment.
- Determine the maximum kinetic energy of photoelectrons ejected by photons of one energy or wavelength, when given the maximum kinetic energy of photoelectrons for a different photon energy or wavelength.

When light strikes materials, it can eject electrons from them. This is called the **photoelectric effect**, meaning that light (*photo*) produces electricity. One common use of the photoelectric effect is in light meters, such as those that adjust the automatic iris on various types of cameras. In a similar way, another use is in solar cells, as you probably have in your calculator or have seen on a roof top or a roadside sign. These make use of the photoelectric effect to convert light into electricity for running different devices.

The photoelectric effect can be observed by allowing light to fall on the metal plate in this evacuated tube. Electrons ejected by the light are collected on the collector wire and

measured as a current. A retarding voltage between the collector wire and plate can then be adjusted so as to determine the energy of the ejected electrons. For example, if it is sufficiently negative, no electrons will reach the wire. (credit: P.P. Urone)

This effect has been known for more than a century and can be studied using a device such as that shown in [link]. This figure shows an evacuated tube with a metal plate and a collector wire that are connected by a variable voltage source, with the collector more negative than the plate. When light (or other EM radiation) strikes the plate in the evacuated tube, it may eject electrons. If the electrons have energy in electron volts (eV) greater than the potential difference between the plate and the wire in volts, some electrons will be collected on the wire. Since the electron energy in eV is qV, where q is the electron charge and V is the potential difference, the electron energy can be measured by adjusting the retarding voltage between the wire and the plate. The voltage that stops the electrons from reaching the wire equals the energy in eV. For example, if $-3.00\ \mathrm{V}$ barely stops the electrons,

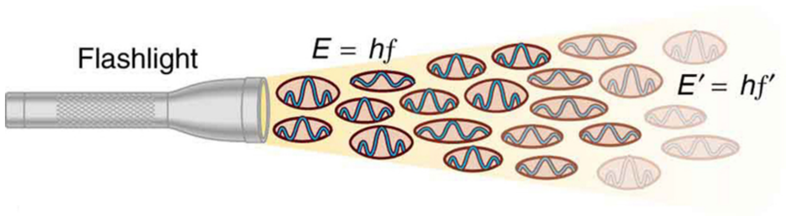
their energy is 3.00 eV. The number of electrons ejected can be determined by measuring the current between the wire and plate. The more light, the more electrons; a little circuitry allows this device to be used as a light meter.

What is really important about the photoelectric effect is what Albert Einstein deduced from it. Einstein realized that there were several characteristics of the photoelectric effect that could be explained only if *EM radiation is itself quantized*: the apparently continuous stream of energy in an EM wave is actually composed of energy quanta called photons. In his explanation of the photoelectric effect, Einstein defined a quantized unit or quantum of EM energy, which we now call a **photon**, with an energy proportional to the frequency of EM radiation. In equation form, the **photon energy** is

Equation:

$$E = \mathrm{hf},$$

where E is the energy of a photon of frequency f and h is Planck's constant. This revolutionary idea looks similar to Planck's quantization of energy states in blackbody oscillators, but it is quite different. It is the quantization of EM radiation itself. EM waves are composed of photons and are not continuous smooth waves as described in previous chapters on optics. Their energy is absorbed and emitted in lumps, not continuously. This is exactly consistent with Planck's quantization of energy levels in blackbody oscillators, since these oscillators increase and decrease their energy in steps of hf by absorbing and emitting photons having $E = \mathrm{hf}$. We do not observe this with our eyes, because there are so many photons in common light sources that individual photons go unnoticed. (See [link].) The next section of the text (Photon Energies and the Electromagnetic Spectrum) is devoted to a discussion of photons and some of their characteristics and implications. For now, we will use the photon concept to explain the photoelectric effect, much as Einstein did.



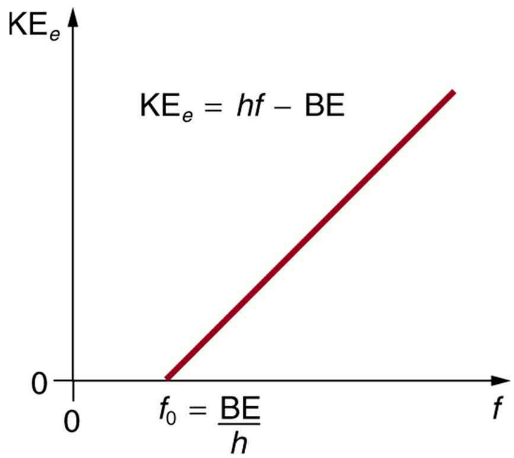
An EM wave of frequency f is composed of photons, or individual quanta of EM radiation. The energy of each photon is $E = \mathrm{hf}$, where h is Planck's constant and f is the frequency of the EM radiation. Higher intensity means more photons per unit area. The flashlight emits large numbers of photons of many different frequencies, hence others have energy $E\prime = \mathrm{hf}\prime$, and so on.

The photoelectric effect has the properties discussed below. All these properties are consistent with the idea that individual photons of EM radiation are absorbed by individual electrons in a material, with the electron gaining the photon's energy. Some of these properties are inconsistent with the idea that EM radiation is a simple wave. For simplicity, let us consider what happens with monochromatic EM radiation in which all photons have the same energy hf.

1. If we vary the frequency of the EM radiation falling on a material, we find the following: For a given material, there is a threshold frequency f_0 for the EM radiation below which no electrons are ejected, regardless of intensity. Individual photons interact with individual electrons. Thus if the photon energy is too small to break an electron away, no electrons will be ejected. If EM radiation was a simple wave, sufficient energy could be obtained by increasing the intensity.
2. *Once EM radiation falls on a material, electrons are ejected without delay*. As soon as an individual photon of a sufficiently high frequency is absorbed by an individual electron, the electron is ejected. If the EM

radiation were a simple wave, several minutes would be required for sufficient energy to be deposited to the metal surface to eject an electron.

3. The number of electrons ejected per unit time is proportional to the intensity of the EM radiation and to no other characteristic. High-intensity EM radiation consists of large numbers of photons per unit area, with all photons having the same characteristic energy hf.
4. If we vary the intensity of the EM radiation and measure the energy of ejected electrons, we find the following: *The maximum kinetic energy of ejected electrons is independent of the intensity of the EM radiation*. Since there are so many electrons in a material, it is extremely unlikely that two photons will interact with the same electron at the same time, thereby increasing the energy given it. Instead (as noted in 3 above), increased intensity results in more electrons of the same energy being ejected. If EM radiation were a simple wave, a higher intensity could give more energy, and higher-energy electrons would be ejected.
5. The kinetic energy of an ejected electron equals the photon energy minus the binding energy of the electron in the specific material. An individual photon can give all of its energy to an electron. The photon's energy is partly used to break the electron away from the material. The remainder goes into the ejected electron's kinetic energy. In equation form, this is given by
   Equation:
   $$\text{KE}_e = \text{hf} - \text{BE},$$
   where KE_e is the maximum kinetic energy of the ejected electron, hf is the photon's energy, and BE is the **binding energy** of the electron to the particular material. (BE is sometimes called the *work function* of the material.) This equation, due to Einstein in 1905, explains the properties of the photoelectric effect quantitatively. An individual photon of EM radiation (it does not come any other way) interacts with an individual electron, supplying enough energy, BE, to break it away, with the remainder going to kinetic energy. The binding energy is $\text{BE} = hf_0$, where f_0 is the threshold frequency for the particular material. [link] shows a graph of maximum KE_e versus the frequency of incident EM radiation falling on a particular material.



Photoelectric effect. A graph of the kinetic energy of an ejected electron, KE_e, versus the frequency of EM radiation impinging on a certain material. There is a threshold frequency below which no electrons are ejected, because the individual photon interacting with an individual electron has insufficient energy to break it away. Above the threshold energy, KE_e increases linearly with f, consistent with $\mathrm{KE}_e = \mathrm{hf} - \mathrm{BE}$. The slope of this line is h — the data can be used to determine Planck's constant experimentally. Einstein gave the first successful explanation of such data by proposing

the idea of photons—
quanta of EM radiation.

Einstein’s idea that EM radiation is quantized was crucial to the beginnings of quantum mechanics. It is a far more general concept than its explanation of the photoelectric effect might imply. All EM radiation can also be modeled in the form of photons, and the characteristics of EM radiation are entirely consistent with this fact. (As we will see in the next section, many aspects of EM radiation, such as the hazards of ultraviolet (UV) radiation, can be explained *only* by photon properties.) More famous for modern relativity, Einstein planted an important seed for quantum mechanics in 1905, the same year he published his first paper on special relativity. His explanation of the photoelectric effect was the basis for the Nobel Prize awarded to him in 1921. Although his other contributions to theoretical physics were also noted in that award, special and general relativity were not fully recognized in spite of having been partially verified by experiment by 1921. Although hero-worshipped, this great man never received Nobel recognition for his most famous work—relativity.

Example:
Calculating Photon Energy and the Photoelectric Effect: A Violet Light

(a) What is the energy in joules and electron volts of a photon of 420-nm violet light? (b) What is the maximum kinetic energy of electrons ejected from calcium by 420-nm violet light, given that the binding energy (or work function) of electrons for calcium metal is 2.71 eV?

Strategy

To solve part (a), note that the energy of a photon is given by $E = \text{hf}$. For part (b), once the energy of the photon is calculated, it is a straightforward application of $\text{KE}_e = \text{hf}–\text{BE}$ to find the ejected electron’s maximum kinetic energy, since BE is given.

Solution for (a)

Photon energy is given by

Equation:

$$E=\text{hf}$$

Since we are given the wavelength rather than the frequency, we solve the familiar relationship $c=f\lambda$ for the frequency, yielding

Equation:

$$f=\frac{c}{\lambda}.$$

Combining these two equations gives the useful relationship

Equation:

$$E=\frac{\text{hc}}{\lambda}.$$

Now substituting known values yields

Equation:

$$E=\frac{\left(6.63\times 10^{-34}\text{ J}\cdot\text{s}\right)\left(3.00\times 10^{8}\text{ m/s}\right)}{420\times 10^{-9}\text{ m}}=4.74\times 10^{-19}\text{ J}.$$

Converting to eV, the energy of the photon is

Equation:

$$E=\left(4.74\times 10^{-19}\text{ J}\right)\frac{1\text{ eV}}{1.6\times 10^{-19}\text{ J}}=2.96\text{ eV}.$$

Solution for (b)

Finding the kinetic energy of the ejected electron is now a simple application of the equation $\text{KE}_e=\text{hf}–\text{BE}$. Substituting the photon energy and binding energy yields

Equation:

$$\text{KE}_e=\text{hf}–\text{BE}=2.96\text{ eV}-2.71\text{ eV}=0.246\text{ eV}.$$

Discussion

The energy of this 420-nm photon of violet light is a tiny fraction of a joule, and so it is no wonder that a single photon would be difficult for us to sense directly—humans are more attuned to energies on the order of joules. But looking at the energy in electron volts, we can see that this photon has enough energy to affect atoms and molecules. A DNA molecule can be broken with about 1 eV of energy, for example, and typical atomic and molecular energies are on the order of eV, so that the UV photon in this example could have biological effects. The ejected electron (called a *photoelectron*) has a rather low energy, and it would not travel far, except in a vacuum. The electron would be stopped by a retarding potential of but 0.26 eV. In fact, if the photon wavelength were longer and its energy less than 2.71 eV, then the formula would give a negative kinetic energy, an impossibility. This simply means that the 420-nm photons with their 2.96-eV energy are not much above the frequency threshold. You can show for yourself that the threshold wavelength is 459 nm (blue light). This means that if calcium metal is used in a light meter, the meter will be insensitive to wavelengths longer than those of blue light. Such a light meter would be completely insensitive to red light, for example.

Note:
PhET Explorations: Photoelectric Effect
See how light knocks electrons off a metal target, and recreate the experiment that spawned the field of quantum mechanics.

https://archive.cnx.org/specials/cf1152da-eae8-11e5-b874-f779884a9994/photoelectric-effect/#sim-photoelectric-effect

Section Summary

- The photoelectric effect is the process in which EM radiation ejects electrons from a material.
- Einstein proposed photons to be quanta of EM radiation having energy $E = \text{hf}$, where f is the frequency of the radiation.

- All EM radiation is composed of photons. As Einstein explained, all characteristics of the photoelectric effect are due to the interaction of individual photons with individual electrons.
- The maximum kinetic energy KE_e of ejected electrons (photoelectrons) is given by $\mathrm{KE}_e = \mathrm{hf} - \mathrm{BE}$, where hf is the photon energy and BE is the binding energy (or work function) of the electron to the particular material.

Conceptual Questions

Exercise:

Problem:

Is visible light the only type of EM radiation that can cause the photoelectric effect?

Exercise:

Problem:

Which aspects of the photoelectric effect cannot be explained without photons? Which can be explained without photons? Are the latter inconsistent with the existence of photons?

Exercise:

Problem:

Is the photoelectric effect a direct consequence of the wave character of EM radiation or of the particle character of EM radiation? Explain briefly.

Exercise:

Problem:

Insulators (nonmetals) have a higher BE than metals, and it is more difficult for photons to eject electrons from insulators. Discuss how this relates to the free charges in metals that make them good conductors.

Exercise:

Problem:

If you pick up and shake a piece of metal that has electrons in it free to move as a current, no electrons fall out. Yet if you heat the metal, electrons can be boiled off. Explain both of these facts as they relate to the amount and distribution of energy involved with shaking the object as compared with heating it.

Problems & Exercises

Exercise:

Problem:

What is the longest-wavelength EM radiation that can eject a photoelectron from silver, given that the binding energy is 4.73 eV? Is this in the visible range?

Solution:

263 nm

Exercise:

Problem:

Find the longest-wavelength photon that can eject an electron from potassium, given that the binding energy is 2.24 eV. Is this visible EM radiation?

Exercise:

Problem:

What is the binding energy in eV of electrons in magnesium, if the longest-wavelength photon that can eject electrons is 337 nm?

Solution:

3.69 eV

Exercise:

Problem:

Calculate the binding energy in eV of electrons in aluminum, if the longest-wavelength photon that can eject them is 304 nm.

Exercise:

Problem:

What is the maximum kinetic energy in eV of electrons ejected from sodium metal by 450-nm EM radiation, given that the binding energy is 2.28 eV?

Solution:

0.483 eV

Exercise:

Problem:

UV radiation having a wavelength of 120 nm falls on gold metal, to which electrons are bound by 4.82 eV. What is the maximum kinetic energy of the ejected photoelectrons?

Exercise:

Problem:

Violet light of wavelength 400 nm ejects electrons with a maximum kinetic energy of 0.860 eV from sodium metal. What is the binding energy of electrons to sodium metal?

Solution:

2.25 eV

Exercise:

Problem:

UV radiation having a 300-nm wavelength falls on uranium metal, ejecting 0.500-eV electrons. What is the binding energy of electrons to uranium metal?

Exercise:

Problem:

What is the wavelength of EM radiation that ejects 2.00-eV electrons from calcium metal, given that the binding energy is 2.71 eV? What type of EM radiation is this?

Solution:

(a) 264 nm

(b) Ultraviolet

Exercise:

Problem:

Find the wavelength of photons that eject 0.100-eV electrons from potassium, given that the binding energy is 2.24 eV. Are these photons visible?

Exercise:

Problem:

What is the maximum velocity of electrons ejected from a material by 80-nm photons, if they are bound to the material by 4.73 eV?

Solution:

$1.95 \times 10^6 \text{ m/s}$

Exercise:

Problem:

Photoelectrons from a material with a binding energy of 2.71 eV are ejected by 420-nm photons. Once ejected, how long does it take these electrons to travel 2.50 cm to a detection device?

Exercise:

Problem:

A laser with a power output of 2.00 mW at a wavelength of 400 nm is projected onto calcium metal. (a) How many electrons per second are ejected? (b) What power is carried away by the electrons, given that the binding energy is 2.71 eV?

Solution:

(a) $4.02 \times 10^{15}\ /\mathrm{s}$

(b) 0.256 mW

Exercise:

Problem:

(a) Calculate the number of photoelectrons per second ejected from a 1.00-mm ^{2} area of sodium metal by 500-nm EM radiation having an intensity of $1.30\ \mathrm{kW/m}^2$ (the intensity of sunlight above the Earth's atmosphere). (b) Given that the binding energy is 2.28 eV, what power is carried away by the electrons? (c) The electrons carry away less power than brought in by the photons. Where does the other power go? How can it be recovered?

Exercise:

Problem: Unreasonable Results

Red light having a wavelength of 700 nm is projected onto magnesium metal to which electrons are bound by 3.68 eV. (a) Use $\mathrm{KE}_e = \mathrm{hf} - \mathrm{BE}$ to calculate the kinetic energy of the ejected electrons.

(b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) –1.90 eV

(b) Negative kinetic energy

(c) That the electrons would be knocked free.

Exercise:

Problem: **Unreasonable Results**

(a) What is the binding energy of electrons to a material from which 4.00-eV electrons are ejected by 400-nm EM radiation? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Glossary

photoelectric effect
: the phenomenon whereby some materials eject electrons when light is shined on them

photon
: a quantum, or particle, of electromagnetic radiation

photon energy
: the amount of energy a photon has; $E = \text{hf}$

binding energy
: also called the *work function*; the amount of energy necessary to eject an electron from a material

Photon Energies and the Electromagnetic Spectrum

- Explain the relationship between the energy of a photon in joules or electron volts and its wavelength or frequency.
- Calculate the number of photons per second emitted by a monochromatic source of specific wavelength and power.

Ionizing Radiation

A photon is a quantum of EM radiation. Its energy is given by $E=\text{hf}$ and is related to the frequency f and wavelength λ of the radiation by
Equation:

$$E=\text{hf}=\frac{\text{hc}}{\lambda}\text{(energy of a photon)},$$

where E is the energy of a single photon and c is the speed of light. When working with small systems, energy in eV is often useful. Note that Planck's constant in these units is
Equation:

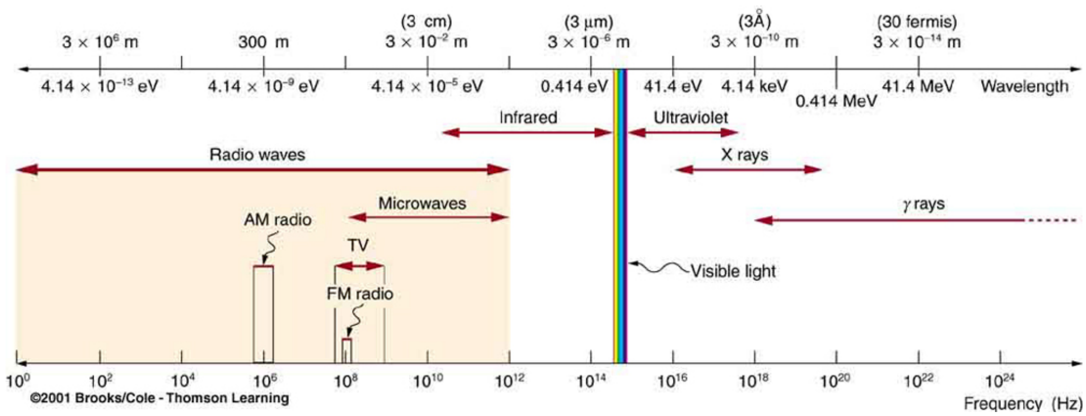
$$h=4.14\times 10^{-15}\text{ eV}\cdot\text{s}.$$

Since many wavelengths are stated in nanometers (nm), it is also useful to know that
Equation:

$$\text{hc}=1240\text{ eV}\cdot\text{nm}.$$

These will make many calculations a little easier.

All EM radiation is composed of photons. [link] shows various divisions of the EM spectrum plotted against wavelength, frequency, and photon energy. Previously in this book, photon characteristics were alluded to in the discussion of some of the characteristics of UV, x rays, and γ rays, the first of which start with frequencies just above violet in the visible spectrum. It was noted that these types of EM radiation have characteristics much different than visible light. We can now see that such properties arise because photon energy is larger at high frequencies.



The EM spectrum, showing major categories as a function of photon energy in eV, as well as wavelength and frequency. Certain characteristics of EM radiation are directly attributable to photon energy alone.

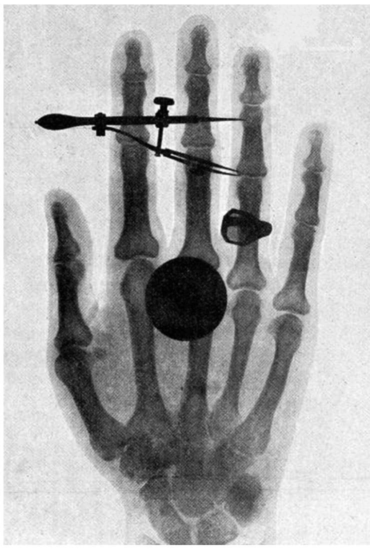
Rotational energies of molecules	10^{-5} eV
Vibrational energies of molecules	0.1 eV
Energy between outer electron shells in atoms	1 eV
Binding energy of a weakly bound molecule	1 eV
Energy of red light	2 eV
Binding energy of a tightly bound molecule	10 eV
Energy to ionize atom or molecule	10 to 1000 eV

Representative Energies for Submicroscopic Effects (Order of Magnitude Only)

Photons act as individual quanta and interact with individual electrons, atoms, molecules, and so on. The energy a photon carries is, thus, crucial to the effects it has. [link] lists representative submicroscopic energies in eV. When we compare photon energies from the EM spectrum in [link] with energies in the table, we can see how effects vary with the type of EM radiation.

Gamma rays, a form of nuclear and cosmic EM radiation, can have the highest frequencies and, hence, the highest photon energies in the EM spectrum. For example, a γ-ray photon with $f = 10^{21}$ Hz has an energy $E = \text{hf} = 6.63 \times 10^{-13}\text{ J} = 4.14\text{ MeV}$. This is sufficient energy to ionize thousands of atoms and molecules, since only 10 to 1000 eV are needed per ionization. In fact, γ rays are one type of **ionizing radiation**, as are x rays and UV, because they produce ionization in materials that absorb

them. Because so much ionization can be produced, a single γ-ray photon can cause significant damage to biological tissue, killing cells or damaging their ability to properly reproduce. When cell reproduction is disrupted, the result can be cancer, one of the known effects of exposure to ionizing radiation. Since cancer cells are rapidly reproducing, they are exceptionally sensitive to the disruption produced by ionizing radiation. This means that ionizing radiation has positive uses in cancer treatment as well as risks in producing cancer.

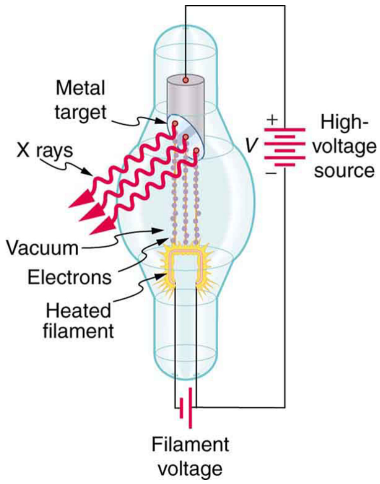


One of the first x-ray images, taken by Röentgen himself. The hand belongs to Bertha Röentgen, his wife. (credit: Wilhelm Conrad Röntgen, via Wikimedia Commons)

High photon energy also enables γ rays to penetrate materials, since a collision with a single atom or molecule is unlikely to absorb all the γ ray's energy. This can make γ rays useful as a probe, and they are sometimes used in medical imaging. **x rays**, as you can see in [link], overlap with the low-frequency end of the γ ray range. Since x rays have energies of keV and up, individual x-ray photons also can produce large amounts of ionization. At lower photon energies, x rays are not as penetrating as γ rays and are slightly less hazardous. X rays are ideal for medical imaging, their most common use, and a fact that was recognized immediately upon their discovery in 1895 by the German physicist W. C. Roentgen (1845–1923). (See [link].) Within one year of their discovery, x rays (for a time called Roentgen rays) were used for medical diagnostics. Roentgen received the 1901 Nobel Prize for the discovery of x rays.

Note:

Connections: Conservation of Energy
Once again, we find that conservation of energy allows us to consider the initial and final forms that energy takes, without having to make detailed calculations of the intermediate steps. [link] is solved by considering only the initial and final forms of energy.



X rays are produced when energetic electrons strike the copper anode of this cathode ray tube (CRT). Electrons (shown here as separate particles) interact individually with the material they strike, sometimes producing photons of EM radiation.

While γ rays originate in nuclear decay, x rays are produced by the process shown in [link]. Electrons ejected by thermal agitation from a hot filament in a vacuum tube are accelerated through a high voltage, gaining kinetic energy from the electrical potential energy. When they strike the anode, the electrons convert their kinetic energy to a variety of forms, including thermal energy. But since an accelerated charge radiates EM waves, and since the electrons act individually, photons are also produced. Some of these x-ray photons obtain the kinetic energy of the electron. The accelerated electrons originate at the cathode, so such a tube is called a cathode ray tube (CRT), and various versions of them are found in older TV and computer screens as well as in x-ray machines.

Example:
X-ray Photon Energy and X-ray Tube Voltage
Find the maximum energy in eV of an x-ray photon produced by electrons accelerated through a potential difference of 50.0 kV in a CRT like the one in [link].
Strategy

Electrons can give all of their kinetic energy to a single photon when they strike the anode of a CRT. (This is something like the photoelectric effect in reverse.) The kinetic energy of the electron comes from electrical potential energy. Thus we can simply equate the maximum photon energy to the electrical potential energy—that is, $hf = qV$. (We do not have to calculate each step from beginning to end if we know that all of the starting energy qV is converted to the final form hf.)

Solution

The maximum photon energy is $\mathrm{hf} = \mathrm{qV}$, where q is the charge of the electron and V is the accelerating voltage. Thus,

Equation:

$$\mathrm{hf} = (1.60 \times 10^{-19}\ \mathrm{C})(50.0 \times 10^{3}\ \mathrm{V}).$$

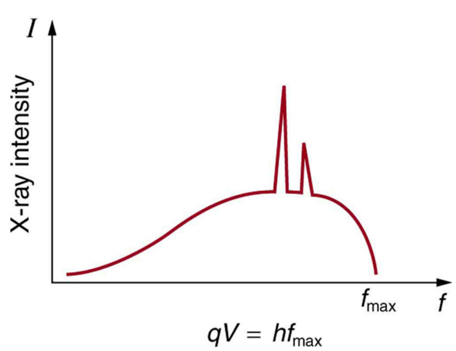
From the definition of the electron volt, we know $1\ \mathrm{eV} = 1.60 \times 10^{-19}\ \mathrm{J}$, where $1\ \mathrm{J} = 1\ \mathrm{C} \cdot \mathrm{V}$. Gathering factors and converting energy to eV yields

Equation:

$$\mathrm{hf} = (50.0 \times 10^{3})(1.60 \times 10^{-19}\ \mathrm{C} \cdot \mathrm{V})\left(\frac{1\ \mathrm{eV}}{1.60 \times 10^{-19}\ \mathrm{C} \cdot \mathrm{V}}\right) = (50.0 \times 10^{3})(1\ \mathrm{eV}) = 50.0\ \mathrm{keV}.$$

Discussion

This example produces a result that can be applied to many similar situations. If you accelerate a single elementary charge, like that of an electron, through a potential given in volts, then its energy in eV has the same numerical value. Thus a 50.0-kV potential generates 50.0 keV electrons, which in turn can produce photons with a maximum energy of 50 keV. Similarly, a 100-kV potential in an x-ray tube can generate up to 100-keV x-ray photons. Many x-ray tubes have adjustable voltages so that various energy x rays with differing energies, and therefore differing abilities to penetrate, can be generated.



X-ray spectrum obtained when energetic electrons strike a material. The smooth part of the spectrum is bremsstrahlung, while the peaks are characteristic of the anode material. Both are atomic processes that produce energetic

photons known as x-ray
photons.

[link] shows the spectrum of x rays obtained from an x-ray tube. There are two distinct features to the spectrum. First, the smooth distribution results from electrons being decelerated in the anode material. A curve like this is obtained by detecting many photons, and it is apparent that the maximum energy is unlikely. This decelerating process produces radiation that is called **bremsstrahlung** (German for *braking radiation*). The second feature is the existence of sharp peaks in the spectrum; these are called **characteristic x rays**, since they are characteristic of the anode material. Characteristic x rays come from atomic excitations unique to a given type of anode material. They are akin to lines in atomic spectra, implying the energy levels of atoms are quantized. Phenomena such as discrete atomic spectra and characteristic x rays are explored further in Atomic Physics.

Ultraviolet radiation (approximately 4 eV to 300 eV) overlaps with the low end of the energy range of x rays, but UV is typically lower in energy. UV comes from the de-excitation of atoms that may be part of a hot solid or gas. These atoms can be given energy that they later release as UV by numerous processes, including electric discharge, nuclear explosion, thermal agitation, and exposure to x rays. A UV photon has sufficient energy to ionize atoms and molecules, which makes its effects different from those of visible light. UV thus has some of the same biological effects as γ rays and x rays. For example, it can cause skin cancer and is used as a sterilizer. The major difference is that several UV photons are required to disrupt cell reproduction or kill a bacterium, whereas single γ-ray and X-ray photons can do the same damage. But since UV does have the energy to alter molecules, it can do what visible light cannot. One of the beneficial aspects of UV is that it triggers the production of vitamin D in the skin, whereas visible light has insufficient energy per photon to alter the molecules that trigger this production. Infantile jaundice is treated by exposing the baby to UV (with eye protection), called phototherapy, the beneficial effects of which are thought to be related to its ability to help prevent the buildup of potentially toxic bilirubin in the blood.

Example:
Photon Energy and Effects for UV
Short-wavelength UV is sometimes called vacuum UV, because it is strongly absorbed by air and must be studied in a vacuum. Calculate the photon energy in eV for 100-nm vacuum UV, and estimate the number of molecules it could ionize or break apart.
Strategy
Using the equation $E = \text{hf}$ and appropriate constants, we can find the photon energy and compare it with energy information in [link].
Solution
The energy of a photon is given by
Equation:

$$E = \text{hf} = \frac{\text{hc}}{\lambda}.$$

Using $\text{hc} = 1240\ \text{eV} \cdot \text{nm}$, we find that
Equation:

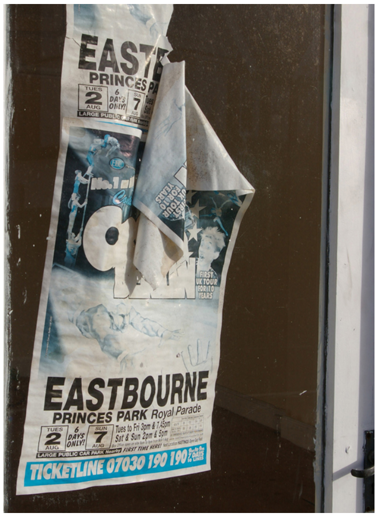
$$E = \frac{\text{hc}}{\lambda} = \frac{1240 \text{ eV} \cdot \text{nm}}{100 \text{ nm}} = 12.4 \text{ eV}.$$

Discussion
According to [link], this photon energy might be able to ionize an atom or molecule, and it is about what is needed to break up a tightly bound molecule, since they are bound by approximately 10 eV. This photon energy could destroy about a dozen weakly bound molecules. Because of its high photon energy, UV disrupts atoms and molecules it interacts with. One good consequence is that all but the longest-wavelength UV is strongly absorbed and is easily blocked by sunglasses. In fact, most of the Sun's UV is absorbed by a thin layer of ozone in the upper atmosphere, protecting sensitive organisms on Earth. Damage to our ozone layer by the addition of such chemicals as CFC's has reduced this protection for us.

Visible Light

The range of photon energies for **visible light** from red to violet is 1.63 to 3.26 eV, respectively (left for this chapter's Problems and Exercises to verify). These energies are on the order of those between outer electron shells in atoms and molecules. This means that these photons can be absorbed by atoms and molecules. A *single* photon can actually stimulate the retina, for example, by altering a receptor molecule that then triggers a nerve impulse. Photons can be absorbed or emitted only by atoms and molecules that have precisely the correct quantized energy step to do so. For example, if a red photon of frequency f encounters a molecule that has an energy step, ΔE, equal to hf, then the photon can be absorbed. Violet flowers absorb red and reflect violet; this implies there is no energy step between levels in the receptor molecule equal to the violet photon's energy, but there is an energy step for the red.

There are some noticeable differences in the characteristics of light between the two ends of the visible spectrum that are due to photon energies. Red light has insufficient photon energy to expose most black-and-white film, and it is thus used to illuminate darkrooms where such film is developed. Since violet light has a higher photon energy, dyes that absorb violet tend to fade more quickly than those that do not. (See [link].) Take a look at some faded color posters in a storefront some time, and you will notice that the blues and violets are the last to fade. This is because other dyes, such as red and green dyes, absorb blue and violet photons, the higher energies of which break up their weakly bound molecules. (Complex molecules such as those in dyes and DNA tend to be weakly bound.) Blue and violet dyes reflect those colors and, therefore, do not absorb these more energetic photons, thus suffering less molecular damage.



Why do the reds, yellows, and greens fade before the blues and violets when exposed to the Sun, as with this poster? The answer is related to photon energy. (credit: Deb Collins, Flickr)

Transparent materials, such as some glasses, do not absorb any visible light, because there is no energy step in the atoms or molecules that could absorb the light. Since individual photons interact with individual atoms, it is nearly impossible to have two photons absorbed simultaneously to reach a large energy step. Because of its lower photon energy, visible light can sometimes pass through many kilometers of a substance, while higher frequencies like UV, x ray, and γ rays are absorbed, because they have sufficient photon energy to ionize the material.

Example:
How Many Photons per Second Does a Typical Light Bulb Produce?
Assuming that 10.0% of a 100-W light bulb's energy output is in the visible range (typical for incandescent bulbs) with an average wavelength of 580 nm, calculate the number of visible photons emitted per second.
Strategy
Power is energy per unit time, and so if we can find the energy per photon, we can determine the number of photons per second. This will best be done in joules, since power is given in watts, which are joules per second.
Solution
The power in visible light production is 10.0% of 100 W, or 10.0 J/s. The energy of the average visible photon is found by substituting the given average wavelength into the formula
Equation:

$$E = \frac{\text{hc}}{\lambda}.$$

This produces

Equation:

$$E = \frac{(6.63 \times 10^{-34}\ \text{J} \cdot \text{s})(3.00 \times 10^{8}\ \text{m/s})}{580 \times 10^{-9}\ \text{m}} = 3.43 \times 10^{-19}\ \text{J}.$$

The number of visible photons per second is thus

Equation:

$$\text{photon/s} = \frac{10.0\ \text{J/s}}{3.43 \times 10^{-19}\ \text{J/photon}} = 2.92 \times 10^{19}\ \text{photon/s}.$$

Discussion

This incredible number of photons per second is verification that individual photons are insignificant in ordinary human experience. It is also a verification of the correspondence principle—on the macroscopic scale, quantization becomes essentially continuous or classical. Finally, there are so many photons emitted by a 100-W lightbulb that it can be seen by the unaided eye many kilometers away.

Lower-Energy Photons

Infrared radiation (IR) has even lower photon energies than visible light and cannot significantly alter atoms and molecules. IR can be absorbed and emitted by atoms and molecules, particularly between closely spaced states. IR is extremely strongly absorbed by water, for example, because water molecules have many states separated by energies on the order of 10^{-5} eV to 10^{-2} eV, well within the IR and microwave energy ranges. This is why in the IR range, skin is almost jet black, with an emissivity near 1—there are many states in water molecules in the skin that can absorb a large range of IR photon energies. Not all molecules have this property. Air, for example, is nearly transparent to many IR frequencies.

Microwaves are the highest frequencies that can be produced by electronic circuits, although they are also produced naturally. Thus microwaves are similar to IR but do not extend to as high frequencies. There are states in water and other molecules that have the same frequency and energy as microwaves, typically about 10^{-5} eV. This is one reason why food absorbs microwaves more strongly than many other materials, making microwave ovens an efficient way of putting energy directly into food.

Photon energies for both IR and microwaves are so low that huge numbers of photons are involved in any significant energy transfer by IR or microwaves (such as warming yourself with a heat lamp or cooking pizza in the microwave). Visible light, IR, microwaves, and all lower frequencies cannot produce ionization with single photons and do not ordinarily have the hazards of higher frequencies. When visible, IR, or microwave radiation *is* hazardous, such as the inducement of cataracts by microwaves, the hazard is due to huge numbers of photons acting together (not to an accumulation of photons, such as sterilization by weak UV). The negative effects of visible, IR, or microwave radiation can be thermal effects, which could be produced by any heat source. But one difference is that at very high intensity, strong electric and magnetic fields can be produced by photons acting together. Such electromagnetic fields (EMF) can actually ionize materials.

Note:
Misconception Alert: High-Voltage Power Lines
Although some people think that living near high-voltage power lines is hazardous to one's health, ongoing studies of the transient field effects produced by these lines show their strengths to be insufficient to cause damage. Demographic studies also fail to show significant correlation of ill effects with high-voltage power lines. The American Physical Society issued a report over 10 years ago on power-line fields, which concluded that the scientific literature and reviews of panels show no consistent, significant link between cancer and power-line fields. They also felt that the "diversion of resources to eliminate a threat which has no persuasive scientific basis is disturbing."

It is virtually impossible to detect individual photons having frequencies below microwave frequencies, because of their low photon energy. But the photons are there. A continuous EM wave can be modeled as photons. At low frequencies, EM waves are generally treated as time- and position-varying electric and magnetic fields with no discernible quantization. This is another example of the correspondence principle in situations involving huge numbers of photons.

Note:
PhET Explorations: Color Vision
Make a whole rainbow by mixing red, green, and blue light. Change the wavelength of a monochromatic beam or filter white light. View the light as a solid beam, or see the individual photons.

https://phet.colorado.edu/sims/html/color-vision/latest/color-vision_en.html

Section Summary

- Photon energy is responsible for many characteristics of EM radiation, being particularly noticeable at high frequencies.
- Photons have both wave and particle characteristics.

Conceptual Questions

Exercise:

Problem: Why are UV, x rays, and γ rays called ionizing radiation?

Exercise:

Problem:

How can treating food with ionizing radiation help keep it from spoiling? UV is not very penetrating. What else could be used?

Exercise:

Problem:

Some television tubes are CRTs. They use an approximately 30-kV accelerating potential to send electrons to the screen, where the electrons stimulate phosphors to emit the light that forms the pictures we watch. Would you expect x rays also to be created?

Exercise:

Problem:

Tanning salons use “safe” UV with a longer wavelength than some of the UV in sunlight. This “safe” UV has enough photon energy to trigger the tanning mechanism. Is it likely to be able to cause cell damage and induce cancer with prolonged exposure?

Exercise:

Problem:

Your pupils dilate when visible light intensity is reduced. Does wearing sunglasses that lack UV blockers increase or decrease the UV hazard to your eyes? Explain.

Exercise:

Problem:

One could feel heat transfer in the form of infrared radiation from a large nuclear bomb detonated in the atmosphere 75 km from you. However, none of the profusely emitted x rays or γ rays reaches you. Explain.

Exercise:

Problem: Can a single microwave photon cause cell damage? Explain.

Exercise:

Problem:

In an x-ray tube, the maximum photon energy is given by $\mathrm{hf} = \mathrm{qV}$. Would it be technically more correct to say $\mathrm{hf} = \mathrm{qV} + \mathrm{BE}$, where BE is the binding energy of electrons in the target anode? Why isn’t the energy stated the latter way?

Problems & Exercises

Exercise:

Problem:

What is the energy in joules and eV of a photon in a radio wave from an AM station that has a 1530-kHz broadcast frequency?

Solution:

6.34×10^{-9} eV, 1.01×10^{-27} J

Exercise:

Problem:

(a) Find the energy in joules and eV of photons in radio waves from an FM station that has a 90.0-MHz broadcast frequency. (b) What does this imply about the number of photons per second that the radio station must broadcast?

Exercise:

Problem: Calculate the frequency in hertz of a 1.00-MeV γ-ray photon.

Solution:

2.42×10^{20} Hz

Exercise:

Problem:

(a) What is the wavelength of a 1.00-eV photon? (b) Find its frequency in hertz. (c) Identify the type of EM radiation.

Exercise:

Problem:

Do the unit conversions necessary to show that $\mathrm{hc} = 1240\ \mathrm{eV} \cdot \mathrm{nm}$, as stated in the text.

Solution:
Equation:

$$\begin{aligned} \mathrm{hc} &= \left(6.62607 \times 10^{-34}\ \mathrm{J} \cdot \mathrm{s}\right)\left(2.99792 \times 10^{8}\ \mathrm{m/s}\right)\left(\frac{10^{9}\ \mathrm{nm}}{1\ \mathrm{m}}\right)\left(\frac{1.00000\ \mathrm{eV}}{1.60218 \times 10^{-19}\ \mathrm{J}}\right) \\ &= 1239.84\ \mathrm{eV} \cdot \mathrm{nm} \\ &\approx 1240\ \mathrm{eV} \cdot \mathrm{nm} \end{aligned}$$

Exercise:

Problem:

Confirm the statement in the text that the range of photon energies for visible light is 1.63 to 3.26 eV, given that the range of visible wavelengths is 380 to 760 nm.

Exercise:

Problem:

(a) Calculate the energy in eV of an IR photon of frequency 2.00×10^{13} Hz. (b) How many of these photons would need to be absorbed simultaneously by a tightly bound molecule to break it apart? (c) What is the energy in eV of a γ ray of frequency 3.00×10^{20} Hz? (d) How many tightly bound molecules could a single such γ ray break apart?

Solution:

(a) 0.0829 eV

(b) 121

(c) 1.24 MeV

(d) 1.24×10^5

Exercise:

Problem: Prove that, to three-digit accuracy, $h = 4.14 \times 10^{-15}$ eV $\cdot$ s, as stated in the text.

Exercise:

Problem:

(a) What is the maximum energy in eV of photons produced in a CRT using a 25.0-kV accelerating potential, such as a color TV? (b) What is their frequency?

Solution:

(a) 25.0×10^3 eV

(b) 6.04×10^{18} Hz

Exercise:

Problem:

What is the accelerating voltage of an x-ray tube that produces x rays with a shortest wavelength of 0.0103 nm?

Exercise:

Problem:

(a) What is the ratio of power outputs by two microwave ovens having frequencies of 950 and 2560 MHz, if they emit the same number of photons per second? (b) What is the ratio of photons per second if they have the same power output?

Solution:

(a) 2.69

(b) 0.371

Exercise:

Problem:

How many photons per second are emitted by the antenna of a microwave oven, if its power output is 1.00 kW at a frequency of 2560 MHz?

Exercise:

Problem:

Some satellites use nuclear power. (a) If such a satellite emits a 1.00-W flux of γ rays having an average energy of 0.500 MeV, how many are emitted per second? (b) These γ rays affect other satellites. How far away must another satellite be to only receive one γ ray per second per square meter?

Solution:

(a) 1.25×10^{13} photons/s

(b) 997 km

Exercise:

Problem:

(a) If the power output of a 650-kHz radio station is 50.0 kW, how many photons per second are produced? (b) If the radio waves are broadcast uniformly in all directions, find the number of photons per second per square meter at a distance of 100 km. Assume no reflection from the ground or absorption by the air.

Exercise:

Problem:

How many x-ray photons per second are created by an x-ray tube that produces a flux of x rays having a power of 1.00 W? Assume the average energy per photon is 75.0 keV.

Solution:

8.33×10^{13} photons/s

Exercise:

Problem:

(a) How far away must you be from a 650-kHz radio station with power 50.0 kW for there to be only one photon per second per square meter? Assume no reflections or absorption, as if you were in deep outer space. (b) Discuss the implications for detecting intelligent life in other solar systems by detecting their radio broadcasts.

Exercise:

Problem:

Assuming that 10.0% of a 100-W light bulb's energy output is in the visible range (typical for incandescent bulbs) with an average wavelength of 580 nm, and that the photons spread out uniformly and are not absorbed by the atmosphere, how far away would you be if 500 photons per second enter the 3.00-mm diameter pupil of your eye? (This number easily stimulates the retina.)

Solution:

181 km

Exercise:

Problem: **Construct Your Own Problem**

Consider a laser pen. Construct a problem in which you calculate the number of photons per second emitted by the pen. Among the things to be considered are the laser pen’s wavelength and power output. Your instructor may also wish for you to determine the minimum diffraction spreading in the beam and the number of photons per square centimeter the pen can project at some large distance. In this latter case, you will also need to consider the output size of the laser beam, the distance to the object being illuminated, and any absorption or scattering along the way.

Glossary

gamma ray
: also γ-ray; highest-energy photon in the EM spectrum

ionizing radiation
: radiation that ionizes materials that absorb it

x ray
: EM photon between γ-ray and UV in energy

bremsstrahlung
: German for *braking radiation*; produced when electrons are decelerated

characteristic x rays
: x rays whose energy depends on the material they were produced in

ultraviolet radiation
: UV; ionizing photons slightly more energetic than violet light

visible light
: the range of photon energies the human eye can detect

infrared radiation
: photons with energies slightly less than red light

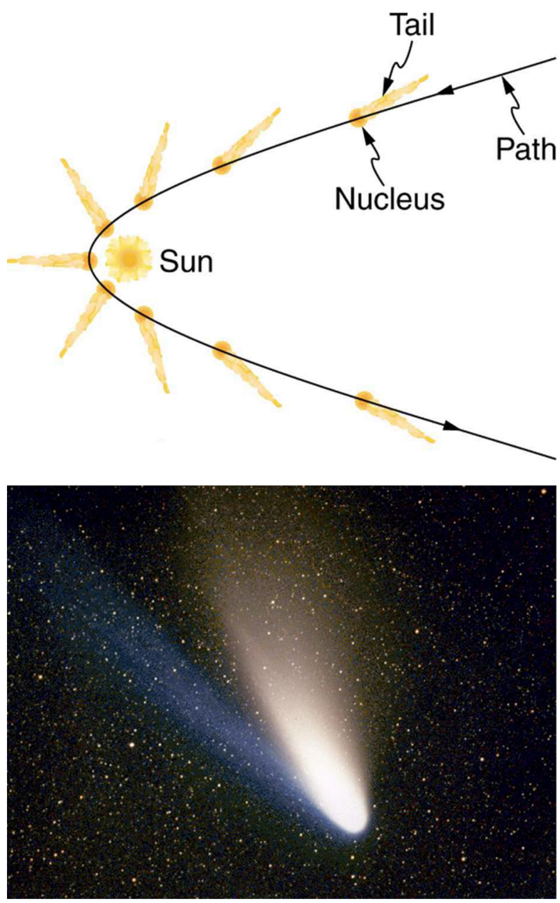
microwaves
: photons with wavelengths on the order of a micron (μm)

Photon Momentum

- Relate the linear momentum of a photon to its energy or wavelength, and apply linear momentum conservation to simple processes involving the emission, absorption, or reflection of photons.
- Account qualitatively for the increase of photon wavelength that is observed, and explain the significance of the Compton wavelength.

Measuring Photon Momentum

The quantum of EM radiation we call a **photon** has properties analogous to those of particles we can see, such as grains of sand. A photon interacts as a unit in collisions or when absorbed, rather than as an extensive wave. Massive quanta, like electrons, also act like macroscopic particles—something we expect, because they are the smallest units of matter. Particles carry momentum as well as energy. Despite photons having no mass, there has long been evidence that EM radiation carries momentum. (Maxwell and others who studied EM waves predicted that they would carry momentum.) It is now a well-established fact that photons *do* have momentum. In fact, photon momentum is suggested by the photoelectric effect, where photons knock electrons out of a substance. [link] shows macroscopic evidence of photon momentum.



The tails of the Hale-Bopp comet point away from the Sun, evidence that light has momentum. Dust emanating from the body of the comet forms this tail. Particles of dust are pushed away from the Sun by light reflecting from them. The blue ionized gas tail is also produced by photons interacting with atoms in the comet material. (credit: Geoff Chester, U.S. Navy, via Wikimedia Commons)

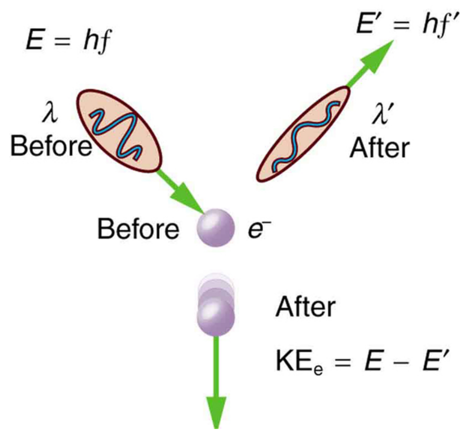
[link] shows a comet with two prominent tails. What most people do not know about the tails is that they always point *away* from the Sun rather than trailing behind the comet (like the tail of Bo Peep's sheep). Comet tails are composed of gases and dust evaporated from the body of the comet and ionized gas. The dust particles recoil away from the Sun when photons scatter from them. Evidently, photons carry momentum in the direction of their motion (away from the Sun), and some of this momentum is transferred to dust particles in collisions. Gas atoms and molecules in the blue tail are most affected by other particles of radiation, such as protons and electrons emanating from the Sun, rather than by the momentum of photons.

Note:
Connections: Conservation of Momentum
Not only is momentum conserved in all realms of physics, but all types of particles are found to have momentum. We expect particles with mass to have momentum, but now we see that massless particles including photons also carry momentum.

Momentum is conserved in quantum mechanics just as it is in relativity and classical physics. Some of the earliest direct experimental evidence of this came from scattering of x-ray photons by electrons in substances, named Compton scattering after the American physicist Arthur H. Compton (1892–1962). Around 1923, Compton observed that x rays scattered from materials had a decreased energy and correctly analyzed this as being due to the scattering of photons from electrons. This phenomenon could be handled as a collision between two particles—a photon and an electron at rest in the material. Energy and momentum are conserved in the collision. (See [link]) He won a Nobel Prize in 1929 for the discovery of this scattering, now called the **Compton effect**, because it helped prove that **photon momentum** is given by
Equation:

$$p = \frac{h}{\lambda},$$

where h is Planck's constant and λ is the photon wavelength. (Note that relativistic momentum given as $p = \gamma m u$ is valid only for particles having mass.)



The Compton effect is the name given to the scattering of a photon by an electron. Energy and momentum are conserved, resulting in a reduction of both for the scattered photon. Studying this effect, Compton verified that photons have momentum.

We can see that photon momentum is small, since $p = h/\lambda$ and h is very small. It is for this reason that we do not ordinarily observe photon

momentum. Our mirrors do not recoil when light reflects from them (except perhaps in cartoons). Compton saw the effects of photon momentum because he was observing x rays, which have a small wavelength and a relatively large momentum, interacting with the lightest of particles, the electron.

Example:

Electron and Photon Momentum Compared

(a) Calculate the momentum of a visible photon that has a wavelength of 500 nm. (b) Find the velocity of an electron having the same momentum. (c) What is the energy of the electron, and how does it compare with the energy of the photon?

Strategy

Finding the photon momentum is a straightforward application of its definition: $p = \frac{h}{\lambda}$. If we find the photon momentum is small, then we can assume that an electron with the same momentum will be nonrelativistic, making it easy to find its velocity and kinetic energy from the classical formulas.

Solution for (a)

Photon momentum is given by the equation:

Equation:

$$p = \frac{h}{\lambda}.$$

Entering the given photon wavelength yields

Equation:

$$p = \frac{6.63 \times 10^{-34}\ \mathrm{J \cdot s}}{500 \times 10^{-9}\ \mathrm{m}} = 1.33 \times 10^{-27}\ \mathrm{kg \cdot m/s}.$$

Solution for (b)

Since this momentum is indeed small, we will use the classical expression $p = \mathrm{mv}$ to find the velocity of an electron with this momentum. Solving for v and using the known value for the mass of an electron gives

Equation:

$$v = \frac{p}{m} = \frac{1.33 \times 10^{-27}\ \text{kg} \cdot \text{m/s}}{9.11 \times 10^{-31}\ \text{kg}} = 1460\ \text{m/s} \approx 1460\ \text{m/s}.$$

Solution for (c)
The electron has kinetic energy, which is classically given by
Equation:

$$\text{KE}_e = \frac{1}{2} m v^2.$$

Thus,
Equation:

$$\text{KE}_e = \frac{1}{2}(9.11 \times 10^{-3}\ \text{kg})(1455\ \text{m/s})^2 = 9.64 \times 10^{-25}\ \text{J}.$$

Converting this to eV by multiplying by $(1\ \text{eV})/(1.602 \times 10^{-19}\ \text{J})$ yields
Equation:

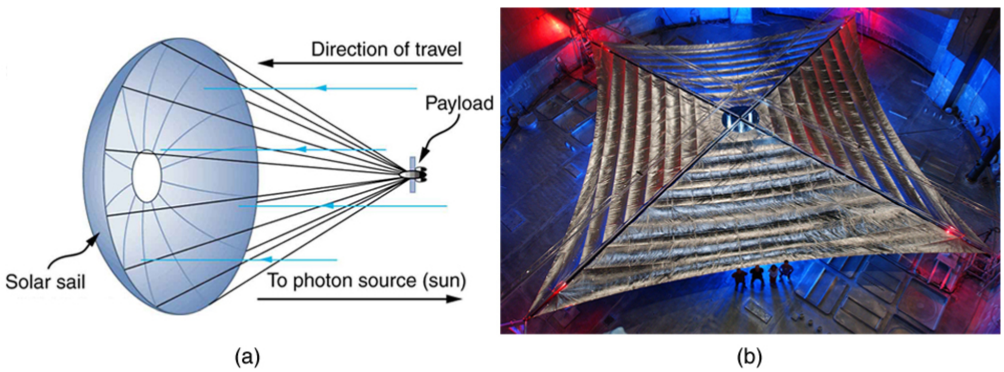
$$\text{KE}_e = 6.02 \times 10^{-6}\ \text{eV}.$$

The photon energy E is
Equation:

$$E = \frac{\text{hc}}{\lambda} = \frac{1240\ \text{eV} \cdot \text{nm}}{500\ \text{nm}} = 2.48\ \text{eV},$$

which is about five orders of magnitude greater.
Discussion
Photon momentum is indeed small. Even if we have huge numbers of them, the total momentum they carry is small. An electron with the same momentum has a 1460 m/s velocity, which is clearly nonrelativistic. A more massive particle with the same momentum would have an even smaller velocity. This is borne out by the fact that it takes far less energy to give an electron the same momentum as a photon. But on a quantum-mechanical scale, especially for high-energy photons interacting with small

masses, photon momentum is significant. Even on a large scale, photon momentum can have an effect if there are enough of them and if there is nothing to prevent the slow recoil of matter. Comet tails are one example, but there are also proposals to build space sails that use huge low-mass mirrors (made of aluminized Mylar) to reflect sunlight. In the vacuum of space, the mirrors would gradually recoil and could actually take spacecraft from place to place in the solar system. (See [link].)



(a) Space sails have been proposed that use the momentum of sunlight reflecting from gigantic low-mass sails to propel spacecraft about the solar system. A Russian test model of this (the Cosmos 1) was launched in 2005, but did not make it into orbit due to a rocket failure. (b) A U.S. version of this, labeled LightSail-1, is scheduled for trial launches in the first part of this decade. It will have a 40-m^2 sail. (credit: Kim Newton/NASA)

Relativistic Photon Momentum

There is a relationship between photon momentum p and photon energy E that is consistent with the relation given previously for the relativistic total energy of a particle as $E^2 = (\text{pc})^2 + (\text{mc})^2$. We know m is zero for a photon, but p is not, so that $E^2 = (\text{pc})^2 + (\text{mc})^2$ becomes

Equation:

$$E=\text{pc},$$

or

Equation:

$$p=\frac{E}{c}\ \text{(photons)}.$$

To check the validity of this relation, note that $E=\text{hc}/\lambda$ for a photon. Substituting this into $p=E/c$ yields

Equation:

$$p=(\text{hc}/\lambda)/c=\frac{h}{\lambda},$$

as determined experimentally and discussed above. Thus, $p=E/c$ is equivalent to Compton's result $p=h/\lambda$. For a further verification of the relationship between photon energy and momentum, see [link].

Note:

Photon Detectors

Almost all detection systems talked about thus far—eyes, photographic plates, photomultiplier tubes in microscopes, and CCD cameras—rely on particle-like properties of photons interacting with a sensitive area. A change is caused and either the change is cascaded or zillions of points are recorded to form an image we detect. These detectors are used in biomedical imaging systems, and there is ongoing research into improving the efficiency of receiving photons, particularly by cooling detection systems and reducing thermal effects.

Example:
Photon Energy and Momentum
Show that $p = E/c$ for the photon considered in the [link].
Strategy
We will take the energy E found in [link], divide it by the speed of light, and see if the same momentum is obtained as before.
Solution
Given that the energy of the photon is 2.48 eV and converting this to joules, we get
Equation:

$$p = \frac{E}{c} = \frac{(2.48 \text{ eV})(1.60 \times 10^{-19} \text{ J/eV})}{3.00 \times 10^{8} \text{ m/s}} = 1.33 \times 10^{-27} \text{ kg} \cdot \text{m/s}.$$

Discussion
This value for momentum is the same as found before (note that unrounded values are used in all calculations to avoid even small rounding errors), an expected verification of the relationship $p = E/c$. This also means the relationship between energy, momentum, and mass given by $E^2 = (\text{pc})^2 + (\text{mc})^2$ applies to both matter and photons. Once again, note that p is not zero, even when m is.

Note:
Problem-Solving Suggestion
Note that the forms of the constants $h = 4.14 \times 10^{-15}\text{ eV} \cdot \text{s}$ and $\text{hc} = 1240\text{ eV} \cdot \text{nm}$ may be particularly useful for this section's Problems and Exercises.

Section Summary

- Photons have momentum, given by $p = \frac{h}{\lambda}$, where λ is the photon wavelength.

- Photon energy and momentum are related by $p = \frac{E}{c}$, where $E = \text{hf} = \text{hc}/\lambda$ for a photon.

Conceptual Questions

Exercise:

Problem:

Which formula may be used for the momentum of all particles, with or without mass?

Exercise:

Problem:

Is there any measurable difference between the momentum of a photon and the momentum of matter?

Exercise:

Problem:

Why don't we feel the momentum of sunlight when we are on the beach?

Problems & Exercises

Exercise:

Problem:

(a) Find the momentum of a 4.00-cm-wavelength microwave photon. (b) Discuss why you expect the answer to (a) to be very small.

Solution:

(a) $1.66 \times 10^{-32}\ \text{kg} \cdot \text{m/s}$

(b) The wavelength of microwave photons is large, so the momentum they carry is very small.

Exercise:

Problem:

(a) What is the momentum of a 0.0100-nm-wavelength photon that could detect details of an atom? (b) What is its energy in MeV?

Exercise:

Problem:

(a) What is the wavelength of a photon that has a momentum of $5.00 \times 10^{-29}\ \text{kg} \cdot \text{m/s}$? (b) Find its energy in eV.

Solution:

(a) 13.3 μm

(b) 9.38×10^{-2} eV

Exercise:

Problem:

(a) A γ-ray photon has a momentum of $8.00 \times 10^{-21}\ \text{kg} \cdot \text{m/s}$. What is its wavelength? (b) Calculate its energy in MeV.

Exercise:

Problem:

(a) Calculate the momentum of a photon having a wavelength of $2.50\ \mu\text{m}$. (b) Find the velocity of an electron having the same momentum. (c) What is the kinetic energy of the electron, and how does it compare with that of the photon?

Solution:

(a) $2.65 \times 10^{-28}\ \text{kg} \cdot \text{m/s}$

(b) 291 m/s

(c) electron 3.86×10^{-26} J, photon 7.96×10^{-20} J, ratio 2.06×10^{6}

Exercise:

Problem:

Repeat the previous problem for a 10.0-nm-wavelength photon.

Exercise:

Problem:

(a) Calculate the wavelength of a photon that has the same momentum as a proton moving at 1.00% of the speed of light. (b) What is the energy of the photon in MeV? (c) What is the kinetic energy of the proton in MeV?

Solution:

(a) 1.32×10^{-13} m

(b) 9.39 MeV

(c) 4.70×10^{-2} MeV

Exercise:

Problem:

(a) Find the momentum of a 100-keV x-ray photon. (b) Find the equivalent velocity of a neutron with the same momentum. (c) What is the neutron's kinetic energy in keV?

Exercise:

Problem:

Take the ratio of relativistic rest energy, $E = \gamma mc^2$, to relativistic momentum, $p = \gamma \mathrm{mu}$, and show that in the limit that mass approaches zero, you find $E/p = c$.

Solution:

$E = \gamma mc^2$ and $P = \gamma mu$, so

Equation:

$$\frac{E}{P} = \frac{\gamma mc^2}{\gamma mu} = \frac{c^2}{u}.$$

As the mass of particle approaches zero, its velocity u will approach c, so that the ratio of energy to momentum in this limit is

Equation:

$$\lim_{m \to 0} \frac{E}{P} = \frac{c^2}{c} = c$$

which is consistent with the equation for photon energy.

Exercise:

Problem: Construct Your Own Problem

Consider a space sail such as mentioned in [link]. Construct a problem in which you calculate the light pressure on the sail in $\mathrm{N/m^2}$ produced by reflecting sunlight. Also calculate the force that could be produced and how much effect that would have on a spacecraft. Among the things to be considered are the intensity of sunlight, its average wavelength, the number of photons per square meter this implies, the area of the space sail, and the mass of the system being accelerated.

Exercise:

Problem: Unreasonable Results

A car feels a small force due to the light it sends out from its headlights, equal to the momentum of the light divided by the time in which it is emitted. (a) Calculate the power of each headlight, if they

exert a total force of 2.00×10^{-2} N backward on the car. (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) 3.00×10^{6} W

(b) Headlights are way too bright.

(c) Force is too large.

Glossary

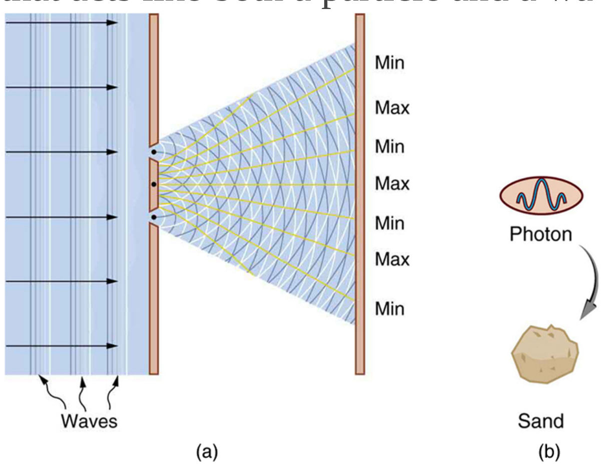
photon momentum
: the amount of momentum a photon has, calculated by $p = \frac{h}{\lambda} = \frac{E}{c}$

Compton effect
: the phenomenon whereby x rays scattered from materials have decreased energy

The Particle-Wave Duality

- Explain what the term particle-wave duality means, and why it is applied to EM radiation.

We have long known that EM radiation is a wave, capable of interference and diffraction. We now see that light can be modeled as photons, which are massless particles. This may seem contradictory, since we ordinarily deal with large objects that never act like both wave and particle. An ocean wave, for example, looks nothing like a rock. To understand small-scale phenomena, we make analogies with the large-scale phenomena we observe directly. When we say something behaves like a wave, we mean it shows interference effects analogous to those seen in overlapping water waves. (See [link].) Two examples of waves are sound and EM radiation. When we say something behaves like a particle, we mean that it interacts as a discrete unit with no interference effects. Examples of particles include electrons, atoms, and photons of EM radiation. How do we talk about a phenomenon that acts like both a particle and a wave?



(a) The interference pattern for light through a double slit is a wave property understood by analogy to water waves. (b) The properties of photons having quantized energy and momentum and acting as a concentrated unit are

understood by analogy to macroscopic particles.

There is no doubt that EM radiation interferes and has the properties of wavelength and frequency. There is also no doubt that it behaves as particles—photons with discrete energy. We call this twofold nature the **particle-wave duality**, meaning that EM radiation has both particle and wave properties. This so-called duality is simply a term for properties of the photon analogous to phenomena we can observe directly, on a macroscopic scale. If this term seems strange, it is because we do not ordinarily observe details on the quantum level directly, and our observations yield either particle *or* wavelike properties, but never both simultaneously.

Since we have a particle-wave duality for photons, and since we have seen connections between photons and matter in that both have momentum, it is reasonable to ask whether there is a particle-wave duality for matter as well. If the EM radiation we once thought to be a pure wave has particle properties, is it possible that matter has wave properties? The answer is yes. The consequences are tremendous, as we will begin to see in the next section.

Note:
PhET Explorations: Quantum Wave Interference
When do photons, electrons, and atoms behave like particles and when do they behave like waves? Watch waves spread out and interfere as they pass through a double slit, then get detected on a screen as tiny dots. Use quantum detectors to explore how measurements change the waves and the patterns they produce on the screen.

Quantum Wave Interference

Section Summary

- EM radiation can behave like either a particle or a wave.
- This is termed particle-wave duality.

Glossary

particle-wave duality
: the property of behaving like either a particle or a wave; the term for the phenomenon that all particles have wave characteristics

The Wave Nature of Matter

- Describe the Davisson-Germer experiment, and explain how it provides evidence for the wave nature of electrons.

De Broglie Wavelength

In 1923 a French physics graduate student named Prince Louis-Victor de Broglie (1892–1987) made a radical proposal based on the hope that nature is symmetric. If EM radiation has both particle and wave properties, then nature would be symmetric if matter also had both particle and wave properties. If what we once thought of as an unequivocal wave (EM radiation) is also a particle, then what we think of as an unequivocal particle (matter) may also be a wave. De Broglie's suggestion, made as part of his doctoral thesis, was so radical that it was greeted with some skepticism. A copy of his thesis was sent to Einstein, who said it was not only probably correct, but that it might be of fundamental importance. With the support of Einstein and a few other prominent physicists, de Broglie was awarded his doctorate.

De Broglie took both relativity and quantum mechanics into account to develop the proposal that *all particles have a wavelength*, given by
Equation:

$$\lambda = \frac{h}{p} \text{ (matter and photons)},$$

where h is Planck's constant and p is momentum. This is defined to be the **de Broglie wavelength**. (Note that we already have this for photons, from the equation $p = h/\lambda$.) The hallmark of a wave is interference. If matter is a wave, then it must exhibit constructive and destructive interference. Why isn't this ordinarily observed? The answer is that in order to see significant interference effects, a wave must interact with an object about the same size as its wavelength. Since h is very small, λ is also small, especially for macroscopic objects. A 3-kg bowling ball moving at 10 m/s, for example, has

Equation:

$$\lambda = h/p = (6.63 \times 10^{-34}\ \mathrm{J{\cdot}s})/[(3\ \mathrm{kg})(10\ \mathrm{m/s}\ )] = 2 \times 10^{-35}\ \mathrm{m}.$$

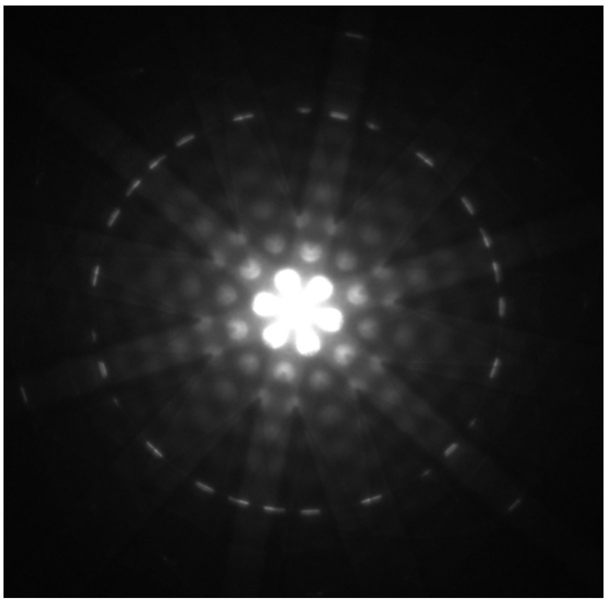
This means that to see its wave characteristics, the bowling ball would have to interact with something about 10^{-35} m in size—far smaller than anything known. When waves interact with objects much larger than their wavelength, they show negligible interference effects and move in straight lines (such as light rays in geometric optics). To get easily observed interference effects from particles of matter, the longest wavelength and hence smallest mass possible would be useful. Therefore, this effect was first observed with electrons.

American physicists Clinton J. Davisson and Lester H. Germer in 1925 and, independently, British physicist G. P. Thomson (son of J. J. Thomson, discoverer of the electron) in 1926 scattered electrons from crystals and found diffraction patterns. These patterns are exactly consistent with interference of electrons having the de Broglie wavelength and are somewhat analogous to light interacting with a diffraction grating. (See [link].)

Note:
Connections: Waves
All microscopic particles, whether massless, like photons, or having mass, like electrons, have wave properties. The relationship between momentum and wavelength is fundamental for all particles.

De Broglie's proposal of a wave nature for all particles initiated a remarkably productive era in which the foundations for quantum mechanics were laid. In 1926, the Austrian physicist Erwin Schrödinger (1887–1961) published four papers in which the wave nature of particles was treated explicitly with wave equations. At the same time, many others began important work. Among them was German physicist Werner Heisenberg

(1901–1976) who, among many other contributions to quantum mechanics, formulated a mathematical treatment of the wave nature of matter that used matrices rather than wave equations. We will deal with some specifics in later sections, but it is worth noting that de Broglie’s work was a watershed for the development of quantum mechanics. De Broglie was awarded the Nobel Prize in 1929 for his vision, as were Davisson and G. P. Thomson in 1937 for their experimental verification of de Broglie’s hypothesis.



This diffraction pattern was obtained for electrons diffracted by crystalline silicon. Bright regions are those of constructive interference, while dark regions are those of destructive interference. (credit: Ndthe, Wikimedia Commons)

Example:

Electron Wavelength versus Velocity and Energy

For an electron having a de Broglie wavelength of 0.167 nm (appropriate for interacting with crystal lattice structures that are about this size): (a) Calculate the electron's velocity, assuming it is nonrelativistic. (b) Calculate the electron's kinetic energy in eV.

Strategy

For part (a), since the de Broglie wavelength is given, the electron's velocity can be obtained from $\lambda = h/p$ by using the nonrelativistic formula for momentum, $p = \mathrm{mv}$. For part (b), once v is obtained (and it has been verified that v is nonrelativistic), the classical kinetic energy is simply $(1/2)mv^2$.

Solution for (a)

Substituting the nonrelativistic formula for momentum ($p = \mathrm{mv}$) into the de Broglie wavelength gives

Equation:

$$\lambda = \frac{h}{p} = \frac{h}{\mathrm{mv}}.$$

Solving for v gives

Equation:

$$v = \frac{h}{m\lambda}.$$

Substituting known values yields

Equation:

$$v = \frac{6.63 \times 10^{-34}\ \mathrm{J \cdot s}}{(9.11 \times 10^{-31}\ \mathrm{kg})(0.167 \times 10^{-9}\ \mathrm{m})} = 4.36 \times 10^{6}\ \mathrm{m/s}.$$

Solution for (b)

While fast compared with a car, this electron's speed is not highly relativistic, and so we can comfortably use the classical formula to find the electron's kinetic energy and convert it to eV as requested.

Equation:

$$\begin{aligned}
\text{KE} &= \frac{1}{2}mv^2 \\
&= \frac{1}{2}(9.11 \times 10^{-31}\ \text{kg})(4.36 \times 10^{6}\ \text{m/s})^2 \\
&= (86.4 \times 10^{-18}\ \text{J})\left(\frac{1\ \text{eV}}{1.602 \times 10^{-19}\ \text{J}}\right) \\
&= 54.0\ \text{eV}
\end{aligned}$$

Discussion

This low energy means that these 0.167-nm electrons could be obtained by accelerating them through a 54.0-V electrostatic potential, an easy task. The results also confirm the assumption that the electrons are nonrelativistic, since their velocity is just over 1% of the speed of light and the kinetic energy is about 0.01% of the rest energy of an electron (0.511 MeV). If the electrons had turned out to be relativistic, we would have had to use more involved calculations employing relativistic formulas.

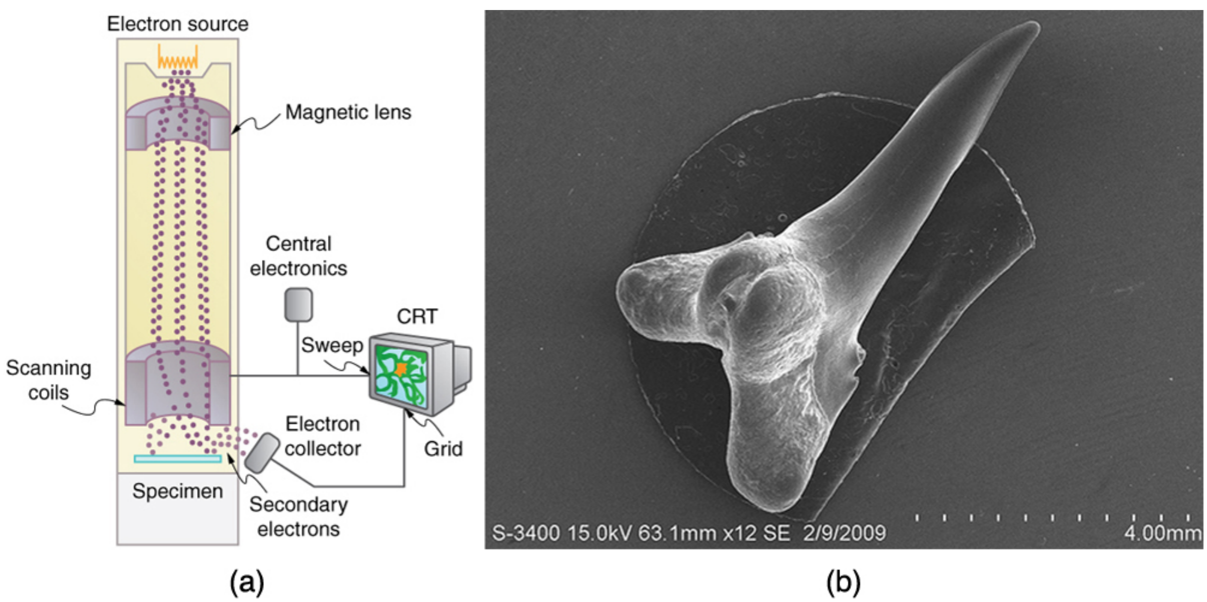
Electron Microscopes

One consequence or use of the wave nature of matter is found in the electron microscope. As we have discussed, there is a limit to the detail observed with any probe having a wavelength. Resolution, or observable detail, is limited to about one wavelength. Since a potential of only 54 V can produce electrons with sub-nanometer wavelengths, it is easy to get electrons with much smaller wavelengths than those of visible light (hundreds of nanometers). Electron microscopes can, thus, be constructed to detect much smaller details than optical microscopes. (See [link].)

There are basically two types of electron microscopes. The transmission electron microscope (TEM) accelerates electrons that are emitted from a hot filament (the cathode). The beam is broadened and then passes through the sample. A magnetic lens focuses the beam image onto a fluorescent screen, a photographic plate, or (most probably) a CCD (light sensitive camera), from which it is transferred to a computer. The TEM is similar to the optical microscope, but it requires a thin sample examined in a vacuum. However it can resolve details as small as 0.1 nm (10^{-10} m), providing magnifications

of 100 million times the size of the original object. The TEM has allowed us to see individual atoms and structure of cell nuclei.

The scanning electron microscope (SEM) provides images by using secondary electrons produced by the primary beam interacting with the surface of the sample (see [link]). The SEM also uses magnetic lenses to focus the beam onto the sample. However, it moves the beam around electrically to "scan" the sample in the *x* and *y* directions. A CCD detector is used to process the data for each electron position, producing images like the one at the beginning of this chapter. The SEM has the advantage of not requiring a thin sample and of providing a 3-D view. However, its resolution is about ten times less than a TEM.



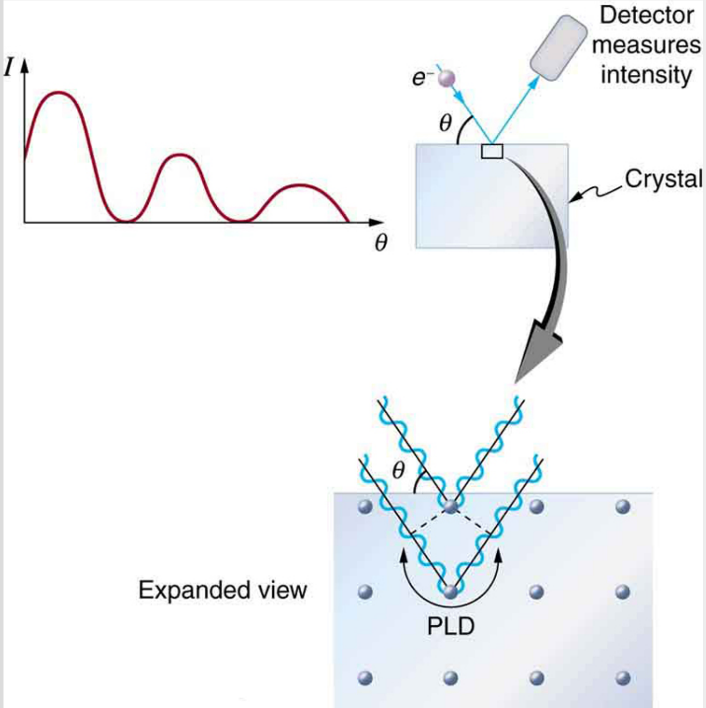
Schematic of a scanning electron microscope (SEM) (a) used to observe small details, such as those seen in this image of a tooth of a *Himipristis*, a type of shark (b). (credit: Dallas Krentzel, Flickr)

Electrons were the first particles with mass to be directly confirmed to have the wavelength proposed by de Broglie. Subsequently, protons, helium nuclei, neutrons, and many others have been observed to exhibit

interference when they interact with objects having sizes similar to their de Broglie wavelength. The de Broglie wavelength for massless particles was well established in the 1920s for photons, and it has since been observed that all massless particles have a de Broglie wavelength $\lambda = h/p.$ The wave nature of all particles is a universal characteristic of nature. We shall see in following sections that implications of the de Broglie wavelength include the quantization of energy in atoms and molecules, and an alteration of our basic view of nature on the microscopic scale. The next section, for example, shows that there are limits to the precision with which we may make predictions, regardless of how hard we try. There are even limits to the precision with which we may measure an object's location or energy.

Note:
Making Connections: A Submicroscopic Diffraction Grating
The wave nature of matter allows it to exhibit all the characteristics of other, more familiar, waves. Diffraction gratings, for example, produce diffraction patterns for light that depend on grating spacing and the wavelength of the light. This effect, as with most wave phenomena, is most pronounced when the wave interacts with objects having a size similar to its wavelength. For gratings, this is the spacing between multiple slits.) When electrons interact with a system having a spacing similar to the electron wavelength, they show the same types of interference patterns as light does for diffraction gratings, as shown at top left in [link].
Atoms are spaced at regular intervals in a crystal as parallel planes, as shown in the bottom part of [link]. The spacings between these planes act like the openings in a diffraction grating. At certain incident angles, the paths of electrons scattering from successive planes differ by one wavelength and, thus, interfere constructively. At other angles, the path length differences are not an integral wavelength, and there is partial to total destructive interference. This type of scattering from a large crystal with well-defined lattice planes can produce dramatic interference patterns. It is called *Bragg reflection*, for the father-and-son team who first explored and analyzed it in some detail. The expanded view also shows the path-length differences and indicates how these depend on incident angle θ in a

manner similar to the diffraction patterns for x rays reflecting from a crystal.



The diffraction pattern at top left is produced by scattering electrons from a crystal and is graphed as a function of incident angle relative to the regular array of atoms in a crystal, as shown at bottom. Electrons scattering from the second layer of atoms travel farther than those scattered from the top layer. If the path length difference (PLD) is an integral wavelength, there is constructive interference.

Let us take the spacing between parallel planes of atoms in the crystal to be d. As mentioned, if the path length difference (PLD) for the electrons is a whole number of wavelengths, there will be constructive interference—that is, $\mathrm{PLD} = n\lambda (n = 1, 2, 3, \ldots)$. Because $\mathrm{AB} = \mathrm{BC} = d \sin \theta$, we have constructive interference when $n\lambda = 2d \sin \theta$. This relationship is

called the *Bragg equation* and applies not only to electrons but also to x rays.
The wavelength of matter is a submicroscopic characteristic that explains a macroscopic phenomenon such as Bragg reflection. Similarly, the wavelength of light is a submicroscopic characteristic that explains the macroscopic phenomenon of diffraction patterns.

Section Summary

- Particles of matter also have a wavelength, called the de Broglie wavelength, given by $\lambda = \frac{h}{p}$, where p is momentum.
- Matter is found to have the same *interference characteristics* as any other wave.

Conceptual Questions

Exercise:

Problem:

How does the interference of water waves differ from the interference of electrons? How are they analogous?

Exercise:

Problem: Describe one type of evidence for the wave nature of matter.

Exercise:

Problem:

Describe one type of evidence for the particle nature of EM radiation.

Problems & Exercises

Exercise:

Problem:

At what velocity will an electron have a wavelength of 1.00 m?

Solution:

7.28×10^{-4} m

Exercise:

Problem:

What is the wavelength of an electron moving at 3.00% of the speed of light?

Exercise:

Problem:

At what velocity does a proton have a 6.00-fm wavelength (about the size of a nucleus)? Assume the proton is nonrelativistic. (1 femtometer = 10^{-15} m.)

Solution:

$6.62 \times 10^{7}\ \mathrm{m/s}$

Exercise:

Problem:

What is the velocity of a 0.400-kg billiard ball if its wavelength is 7.50 cm (large enough for it to interfere with other billiard balls)?

Exercise:

Problem:

Find the wavelength of a proton moving at 1.00% of the speed of light.

Solution:

1.32×10^{-13} m

Exercise:

Problem:

Experiments are performed with ultracold neutrons having velocities as small as 1.00 m/s. (a) What is the wavelength of such a neutron? (b) What is its kinetic energy in eV?

Exercise:

Problem:

(a) Find the velocity of a neutron that has a 6.00-fm wavelength (about the size of a nucleus). Assume the neutron is nonrelativistic. (b) What is the neutron's kinetic energy in MeV?

Solution:

(a) 6.62×10^{7} m/s

(b) 22.9 MeV

Exercise:

Problem:

What is the wavelength of an electron accelerated through a 30.0-kV potential, as in a TV tube?

Exercise:

Problem:

What is the kinetic energy of an electron in a TEM having a 0.0100-nm wavelength?

Solution:
Equation: 15.1 keV

Exercise:

Problem:

(a) Calculate the velocity of an electron that has a wavelength of $1.00\ \mu\text{m}$. (b) Through what voltage must the electron be accelerated to have this velocity?

Exercise:

Problem:

The velocity of a proton emerging from a Van de Graaff accelerator is 25.0% of the speed of light. (a) What is the proton's wavelength? (b) What is its kinetic energy, assuming it is nonrelativistic? (c) What was the equivalent voltage through which it was accelerated?

Solution:

(a) 5.29 fm

(b) $4.70 \times 10^{-12}\ \text{J}$

(c) 29.4 MV

Exercise:

Problem:

The kinetic energy of an electron accelerated in an x-ray tube is 100 keV. Assuming it is nonrelativistic, what is its wavelength?

Exercise:

Problem: Unreasonable Results

(a) Assuming it is nonrelativistic, calculate the velocity of an electron with a 0.100-fm wavelength (small enough to detect details of a nucleus). (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) 7.28×10^{12} m/s

(b) This is thousands of times the speed of light (an impossibility).

(c) The assumption that the electron is non-relativistic is unreasonable at this wavelength.

Glossary

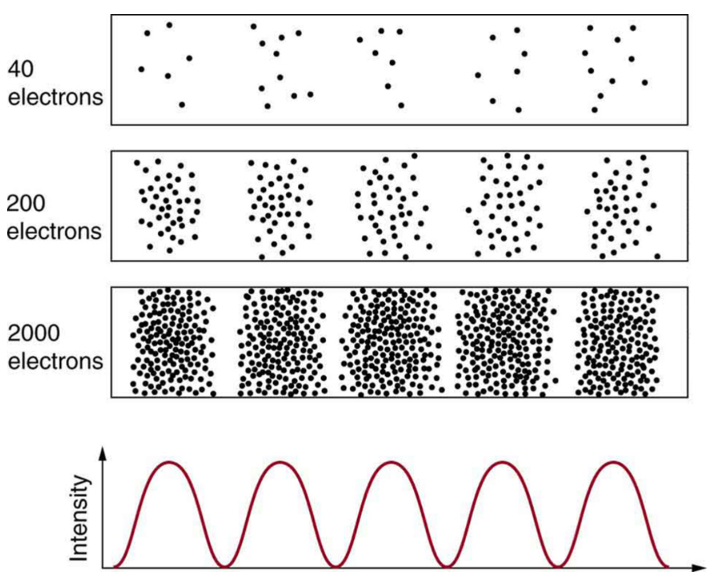
de Broglie wavelength
: the wavelength possessed by a particle of matter, calculated by $\lambda = h/p$

Probability: The Heisenberg Uncertainty Principle

- Use both versions of Heisenberg’s uncertainty principle in calculations.
- Explain the implications of Heisenberg’s uncertainty principle for measurements.

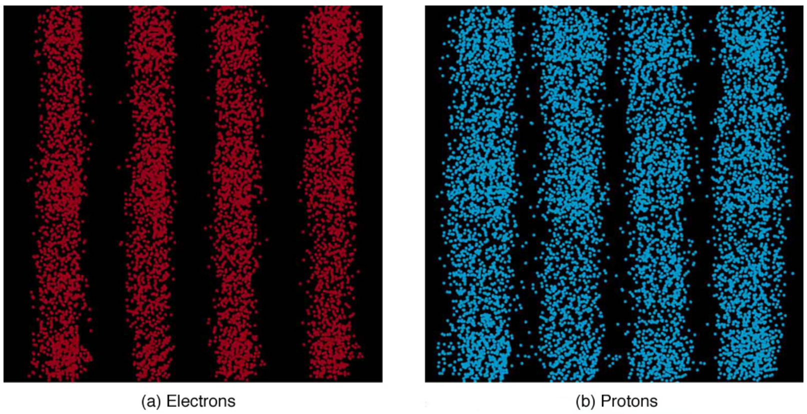
Probability Distribution

Matter and photons are waves, implying they are spread out over some distance. What is the position of a particle, such as an electron? Is it at the center of the wave? The answer lies in how you measure the position of an electron. Experiments show that you will find the electron at some definite location, unlike a wave. But if you set up exactly the same situation and measure it again, you will find the electron in a different location, often far outside any experimental uncertainty in your measurement. Repeated measurements will display a statistical distribution of locations that appears wavelike. (See [link].)



The building up of the diffraction pattern of electrons scattered from a crystal surface. Each electron arrives

at a definite location, which cannot be precisely predicted. The overall distribution shown at the bottom can be predicted as the diffraction of waves having the de Broglie wavelength of the electrons.



Double-slit interference for electrons (a) and protons (b) is identical for equal wavelengths and equal slit separations. Both patterns are probability distributions in the sense that they are built up by individual particles traversing the apparatus, the paths of which are not individually predictable.

After de Broglie proposed the wave nature of matter, many physicists, including Schrödinger and Heisenberg, explored the consequences. The idea quickly emerged that, *because of its wave character, a particle's trajectory and destination cannot be precisely predicted for each particle individually*. However, each particle goes to a definite place (as illustrated in [link]). After compiling enough data, you get a distribution related to the

particle’s wavelength and diffraction pattern. There is a certain *probability* of finding the particle at a given location, and the overall pattern is called a **probability distribution**. Those who developed quantum mechanics devised equations that predicted the probability distribution in various circumstances.

It is somewhat disquieting to think that you cannot predict exactly where an individual particle will go, or even follow it to its destination. Let us explore what happens if we try to follow a particle. Consider the double-slit patterns obtained for electrons and photons in [link]. First, we note that these patterns are identical, following $d \sin \theta = m\lambda$, the equation for double-slit constructive interference developed in Photon Energies and the Electromagnetic Spectrum, where d is the slit separation and λ is the electron or photon wavelength.

Both patterns build up statistically as individual particles fall on the detector. This can be observed for photons or electrons—for now, let us concentrate on electrons. You might imagine that the electrons are interfering with one another as any waves do. To test this, you can lower the intensity until there is never more than one electron between the slits and the screen. The same interference pattern builds up! This implies that a particle’s probability distribution spans both slits, and the particles actually interfere with themselves. Does this also mean that the electron goes through both slits? An electron is a basic unit of matter that is not divisible. But it is a fair question, and so we should look to see if the electron traverses one slit or the other, or both. One possibility is to have coils around the slits that detect charges moving through them. What is observed is that an electron always goes through one slit or the other; it does not split to go through both. But there is a catch. If you determine that the electron went through one of the slits, you no longer get a double slit pattern—instead, you get single slit interference. There is no escape by using another method of determining which slit the electron went through. Knowing the particle went through one slit forces a single-slit pattern. If you do not observe which slit the electron goes through, you obtain a double-slit pattern.

Heisenberg Uncertainty

How does knowing which slit the electron passed through change the pattern? The answer is fundamentally important—*measurement affects the system being observed*. Information can be lost, and in some cases it is impossible to measure two physical quantities simultaneously to exact precision. For example, you can measure the position of a moving electron by scattering light or other electrons from it. Those probes have momentum themselves, and by scattering from the electron, they change its momentum *in a manner that loses information*. There is a limit to absolute knowledge, even in principle.

Werner Heisenberg was one of the best of those physicists who developed early quantum mechanics. Not only did his work enable a description of nature on the very small scale, it also changed our

view of the availability of knowledge. Although he is universally recognized for his brilliance and the importance of his work (he received the Nobel Prize in 1932, for example), Heisenberg remained in Germany during World War II and headed the German effort to build a nuclear bomb, permanently alienating himself from most of the scientific community. (credit: Author Unknown, via Wikimedia Commons)

It was Werner Heisenberg who first stated this limit to knowledge in 1929 as a result of his work on quantum mechanics and the wave characteristics of all particles. (See [link]). Specifically, consider simultaneously measuring the position and momentum of an electron (it could be any particle). There is an **uncertainty in position** Δx that is approximately equal to the wavelength of the particle. That is,

Equation:

$$\Delta x \approx \lambda.$$

As discussed above, a wave is not located at one point in space. If the electron's position is measured repeatedly, a spread in locations will be observed, implying an uncertainty in position Δx. To detect the position of the particle, we must interact with it, such as having it collide with a detector. In the collision, the particle will lose momentum. This change in momentum could be anywhere from close to zero to the total momentum of the particle, $p = h/\lambda$. It is not possible to tell how much momentum will be transferred to a detector, and so there is an **uncertainty in momentum** Δp, too. In fact, the uncertainty in momentum may be as large as the momentum itself, which in equation form means that

Equation:

$$\Delta p \approx \frac{h}{\lambda}.$$

The uncertainty in position can be reduced by using a shorter-wavelength electron, since $\Delta x \approx \lambda$. But shortening the wavelength increases the uncertainty in momentum, since $\Delta p \approx h/\lambda$. Conversely, the uncertainty in momentum can be reduced by using a longer-wavelength electron, but this increases the uncertainty in position. Mathematically, you can express this trade-off by multiplying the uncertainties. The wavelength cancels, leaving

Equation:

$$\Delta x \Delta p \approx h.$$

So if one uncertainty is reduced, the other must increase so that their product is $\approx h$.

With the use of advanced mathematics, Heisenberg showed that the best that can be done in a *simultaneous measurement of position and momentum* is

Equation:

$$\Delta x \Delta p \geq \frac{h}{4\pi}.$$

This is known as the **Heisenberg uncertainty principle**. It is impossible to measure position x and momentum p simultaneously with uncertainties Δx and Δp that multiply to be less than $h/4\pi$. Neither uncertainty can be zero. Neither uncertainty can become small without the other becoming large. A small wavelength allows accurate position measurement, but it increases the momentum of the probe to the point that it further disturbs the momentum of a system being measured. For example, if an electron is scattered from an atom and has a wavelength small enough to detect the position of electrons in the atom, its momentum can knock the electrons from their orbits in a manner that loses information about their original motion. It is therefore impossible to follow an electron in its orbit around an atom. If you measure the electron's position, you will find it in a definite location, but the atom will be disrupted. Repeated measurements on identical atoms will produce interesting probability distributions for electrons around the atom, but they will not produce motion information. The probability distributions are referred to as electron clouds or orbitals. The shapes of these orbitals are often shown in general chemistry texts and are discussed in The Wave Nature of Matter Causes Quantization.

Example:
Heisenberg Uncertainty Principle in Position and Momentum for an Atom

(a) If the position of an electron in an atom is measured to an accuracy of 0.0100 nm, what is the electron's uncertainty in velocity? (b) If the electron has this velocity, what is its kinetic energy in eV?

Strategy

The uncertainty in position is the accuracy of the measurement, or $\Delta x = 0.0100$ nm. Thus the smallest uncertainty in momentum Δp can be calculated using $\Delta x \Delta p \geq h/4\pi$. Once the uncertainty in momentum Δp is found, the uncertainty in velocity can be found from $\Delta p = m\Delta v$.

Solution for (a)

Using the equals sign in the uncertainty principle to express the minimum uncertainty, we have

Equation:

$$\Delta x \Delta p = \frac{h}{4\pi}.$$

Solving for Δp and substituting known values gives

Equation:

$$\Delta p = \frac{h}{4\pi \Delta x} = \frac{6.63 \times 10^{-34}\ \mathrm{J \cdot s}}{4\pi(1.00 \times 10^{-11}\ \mathrm{m})} = 5.28 \times 10^{-24}\ \mathrm{kg \cdot m/s}.$$

Thus,

Equation:

$$\Delta p = 5.28 \times 10^{-24}\ \mathrm{kg \cdot m/s} = m\Delta v.$$

Solving for Δv and substituting the mass of an electron gives

Equation:

$$\Delta v = \frac{\Delta p}{m} = \frac{5.28 \times 10^{-24}\ \mathrm{kg \cdot m/s}}{9.11 \times 10^{-31}\ \mathrm{kg}} = 5.79 \times 10^{6}\ \mathrm{m/s}.$$

Solution for (b)

Although large, this velocity is not highly relativistic, and so the electron's kinetic energy is

Equation:

$$\begin{aligned} \mathrm{KE}_e &= \tfrac{1}{2}mv^2 \\ &= \tfrac{1}{2}(9.11 \times 10^{-31}\ \mathrm{kg})(5.79 \times 10^{6}\ \mathrm{m/s})^2 \\ &= (1.53 \times 10^{-17}\ \mathrm{J})\left(\tfrac{1\ \mathrm{eV}}{1.60 \times 10^{-19}\ \mathrm{J}}\right) = 95.5\ \mathrm{eV}. \end{aligned}$$

Discussion

Since atoms are roughly 0.1 nm in size, knowing the position of an electron to 0.0100 nm localizes it reasonably well inside the atom. This

would be like being able to see details one-tenth the size of the atom. But the consequent uncertainty in velocity is large. You certainly could not follow it very well if its velocity is so uncertain. To get a further idea of how large the uncertainty in velocity is, we assumed the velocity of the electron was equal to its uncertainty and found this gave a kinetic energy of 95.5 eV. This is significantly greater than the typical energy difference between levels in atoms (see [link]), so that it is impossible to get a meaningful energy for the electron if we know its position even moderately well.

Why don't we notice Heisenberg's uncertainty principle in everyday life? The answer is that Planck's constant is very small. Thus the lower limit in the uncertainty of measuring the position and momentum of large objects is negligible. We can detect sunlight reflected from Jupiter and follow the planet in its orbit around the Sun. The reflected sunlight alters the momentum of Jupiter and creates an uncertainty in its momentum, but this is totally negligible compared with Jupiter's huge momentum. The correspondence principle tells us that the predictions of quantum mechanics become indistinguishable from classical physics for large objects, which is the case here.

Heisenberg Uncertainty for Energy and Time

There is another form of **Heisenberg's uncertainty principle** for *simultaneous measurements of energy and time*. In equation form,
Equation:

$$\Delta E \Delta t \geq \frac{h}{4\pi},$$

where ΔE is the **uncertainty in energy** and Δt is the **uncertainty in time**. This means that within a time interval Δt, it is not possible to measure energy precisely—there will be an uncertainty ΔE in the measurement. In order to measure energy more precisely (to make ΔE smaller), we must

increase Δt. This time interval may be the amount of time we take to make the measurement, or it could be the amount of time a particular state exists, as in the next [link].

Example:
Heisenberg Uncertainty Principle for Energy and Time for an Atom
An atom in an excited state temporarily stores energy. If the lifetime of this excited state is measured to be 1.0×10^{-10} s, what is the minimum uncertainty in the energy of the state in eV?
Strategy
The minimum uncertainty in energy ΔE is found by using the equals sign in $\Delta E \Delta t \geq h/4\pi$ and corresponds to a reasonable choice for the uncertainty in time. The largest the uncertainty in time can be is the full lifetime of the excited state, or $\Delta t = 1.0\times10^{-10}$ s.
Solution
Solving the uncertainty principle for ΔE and substituting known values gives
Equation:

$$\Delta E = \frac{h}{4\pi\Delta t} = \frac{6.63 \times 10^{-34}\ \text{J} \cdot \text{s}}{4\pi(1.0\times10^{-10}\ \text{s})} = 5.3 \times 10^{-25}\ \text{J}.$$

Now converting to eV yields
Equation:

$$\Delta E = (5.3 \times 10^{-25}\ \text{J})\left(\frac{1\ \text{eV}}{1.6 \times 10^{-19}\ \text{J}}\right) = 3.3 \times 10^{-6}\ \text{eV}.$$

Discussion
The lifetime of 10^{-10} s is typical of excited states in atoms—on human time scales, they quickly emit their stored energy. An uncertainty in energy of only a few millionths of an eV results. This uncertainty is small compared with typical excitation energies in atoms, which are on the order of 1 eV. So here the uncertainty principle limits the accuracy with which

we can measure the lifetime and energy of such states, but not very significantly.

The uncertainty principle for energy and time can be of great significance if the lifetime of a system is very short. Then Δt is very small, and ΔE is consequently very large. Some nuclei and exotic particles have extremely short lifetimes (as small as 10^{-25} s), causing uncertainties in energy as great as many GeV (10^9 eV). Stored energy appears as increased rest mass, and so this means that there is significant uncertainty in the rest mass of short-lived particles. When measured repeatedly, a spread of masses or decay energies are obtained. The spread is ΔE. You might ask whether this uncertainty in energy could be avoided by not measuring the lifetime. The answer is no. Nature knows the lifetime, and so its brevity affects the energy of the particle. This is so well established experimentally that the uncertainty in decay energy is used to calculate the lifetime of short-lived states. Some nuclei and particles are so short-lived that it is difficult to measure their lifetime. But if their decay energy can be measured, its spread is ΔE, and this is used in the uncertainty principle ($\Delta E \Delta t \geq h/4\pi$) to calculate the lifetime Δt.

There is another consequence of the uncertainty principle for energy and time. If energy is uncertain by ΔE, then conservation of energy can be violated by ΔE for a time Δt. Neither the physicist nor nature can tell that conservation of energy has been violated, if the violation is temporary and smaller than the uncertainty in energy. While this sounds innocuous enough, we shall see in later chapters that it allows the temporary creation of matter from nothing and has implications for how nature transmits forces over very small distances.

Finally, note that in the discussion of particles and waves, we have stated that individual measurements produce precise or particle-like results. A definite position is determined each time we observe an electron, for example. But repeated measurements produce a spread in values consistent with wave characteristics. The great theoretical physicist Richard Feynman (1918–1988) commented, “What there are, are particles.” When you

observe enough of them, they distribute themselves as you would expect for a wave phenomenon. However, what there are as they travel we cannot tell because, when we do try to measure, we affect the traveling.

Section Summary

- Matter is found to have the same interference characteristics as any other wave.
- There is now a probability distribution for the location of a particle rather than a definite position.
- Another consequence of the wave character of all particles is the Heisenberg uncertainty principle, which limits the precision with which certain physical quantities can be known simultaneously. For position and momentum, the uncertainty principle is $\Delta x \Delta p \geq \frac{h}{4\pi}$, where Δx is the uncertainty in position and Δp is the uncertainty in momentum.
- For energy and time, the uncertainty principle is $\Delta E \Delta t \geq \frac{h}{4\pi}$ where ΔE is the uncertainty in energy and Δt is the uncertainty in time.
- These small limits are fundamentally important on the quantum-mechanical scale.

Conceptual Questions

Exercise:

Problem:

What is the Heisenberg uncertainty principle? Does it place limits on what can be known?

Problems & Exercises

Exercise:

Problem:

(a) If the position of an electron in a membrane is measured to an accuracy of $1.00\ \mu\text{m}$, what is the electron's minimum uncertainty in velocity? (b) If the electron has this velocity, what is its kinetic energy in eV? (c) What are the implications of this energy, comparing it to typical molecular binding energies?

Solution:

(a) 57.9 m/s

(b) 9.55×10^{-9} eV

(c) From [link], we see that typical molecular binding energies range from about 1eV to 10 eV, therefore the result in part (b) is approximately 9 orders of magnitude smaller than typical molecular binding energies.

Exercise:

Problem:

(a) If the position of a chlorine ion in a membrane is measured to an accuracy of $1.00\ \mu\text{m}$, what is its minimum uncertainty in velocity, given its mass is 5.86×10^{-26} kg? (b) If the ion has this velocity, what is its kinetic energy in eV, and how does this compare with typical molecular binding energies?

Exercise:

Problem:

Suppose the velocity of an electron in an atom is known to an accuracy of $2.0 \times 10^{3}\ \text{m/s}$ (reasonably accurate compared with orbital velocities). What is the electron's minimum uncertainty in position, and how does this compare with the approximate 0.1-nm size of the atom?

Solution:

29 nm,

290 times greater

Exercise:

Problem:

The velocity of a proton in an accelerator is known to an accuracy of 0.250% of the speed of light. (This could be small compared with its velocity.) What is the smallest possible uncertainty in its position?

Exercise:

Problem:

A relatively long-lived excited state of an atom has a lifetime of 3.00 ms. What is the minimum uncertainty in its energy?

Solution:

1.10×10^{-13} eV

Exercise:

Problem:

(a) The lifetime of a highly unstable nucleus is 10^{-20} s. What is the smallest uncertainty in its decay energy? (b) Compare this with the rest energy of an electron.

Exercise:

Problem:

The decay energy of a short-lived particle has an uncertainty of 1.0 MeV due to its short lifetime. What is the smallest lifetime it can have?

Solution:

3.3×10^{-22} s

Exercise:

Problem:

The decay energy of a short-lived nuclear excited state has an uncertainty of 2.0 eV due to its short lifetime. What is the smallest lifetime it can have?

Exercise:

Problem:

What is the approximate uncertainty in the mass of a muon, as determined from its decay lifetime?

Solution:

2.66×10^{-46} kg

Exercise:

Problem:

Derive the approximate form of Heisenberg's uncertainty principle for energy and time, $\Delta E \Delta t \approx h$, using the following arguments: Since the position of a particle is uncertain by $\Delta x \approx \lambda$, where λ is the wavelength of the photon used to examine it, there is an uncertainty in the time the photon takes to traverse Δx. Furthermore, the photon has an energy related to its wavelength, and it can transfer some or all of this energy to the object being examined. Thus the uncertainty in the energy of the object is also related to λ. Find Δt and ΔE; then multiply them to give the approximate uncertainty principle.

Glossary

Heisenberg's uncertainty principle
: a fundamental limit to the precision with which pairs of quantities (momentum and position, and energy and time) can be measured

uncertainty in energy
: lack of precision or lack of knowledge of precise results in measurements of energy

uncertainty in time
: lack of precision or lack of knowledge of precise results in measurements of time

uncertainty in momentum
: lack of precision or lack of knowledge of precise results in measurements of momentum

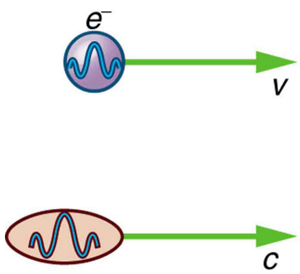
uncertainty in position
: lack of precision or lack of knowledge of precise results in measurements of position

probability distribution
: the overall spatial distribution of probabilities to find a particle at a given location

The Particle-Wave Duality Reviewed

- Explain the concept of particle-wave duality, and its scope.

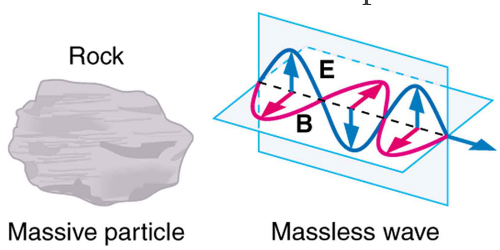
Particle-wave duality—the fact that all particles have wave properties—is one of the cornerstones of quantum mechanics. We first came across it in the treatment of photons, those particles of EM radiation that exhibit both particle and wave properties, but not at the same time. Later it was noted that particles of matter have wave properties as well. The dual properties of particles and waves are found for all particles, whether massless like photons, or having a mass like electrons. (See [link].)



On a quantum-mechanical scale (i.e., very small), particles with and without mass have wave properties. For example, both electrons and photons have wavelengths but also behave as particles.

There are many submicroscopic particles in nature. Most have mass and are expected to act as particles, or the smallest units of matter. All these masses have wave properties, with wavelengths given by the de Broglie relationship $\lambda = h/p$. So, too, do combinations of these particles, such as nuclei, atoms, and molecules. As a combination of masses becomes large, particularly if it is large enough to be called macroscopic, its wave nature becomes difficult to observe. This is consistent with our common experience with matter.

Some particles in nature are massless. We have only treated the photon so far, but all massless entities travel at the speed of light, have a wavelength, and exhibit particle and wave behaviors. They have momentum given by a rearrangement of the de Broglie relationship, $p = h/\lambda$. In large combinations of these massless particles (such large combinations are common only for photons or EM waves), there is mostly wave behavior upon detection, and the particle nature becomes difficult to observe. This is also consistent with experience. (See [link].)



On a classical scale (macroscopic), particles with mass behave as particles and not as waves. Particles without mass act as waves and not as particles.

The particle-wave duality is a universal attribute. It is another connection between matter and energy. Not only has modern physics been able to

describe nature for high speeds and small sizes, it has also discovered new connections and symmetries. There is greater unity and symmetry in nature than was known in the classical era—but they were dreamt of. A beautiful poem written by the English poet William Blake some two centuries ago contains the following four lines:

To see the World in a Grain of Sand

And a Heaven in a Wild Flower

Hold Infinity in the palm of your hand

And Eternity in an hour

Integrated Concepts

The problem set for this section involves concepts from this chapter and several others. Physics is most interesting when applied to general situations involving more than a narrow set of physical principles. For example, photons have momentum, hence the relevance of Linear Momentum and Collisions. The following topics are involved in some or all of the problems in this section:

- Dynamics: Newton's Laws of Motion
- Work, Energy, and Energy Resources
- Linear Momentum and Collisions
- Heat and Heat Transfer Methods
- Electric Potential and Electric Field
- Electric Current, Resistance, and Ohm's Law
- Wave Optics
- Special Relativity

Note:
Problem-Solving Strategy

1. Identify which physical principles are involved.

2. Solve the problem using strategies outlined in the text.

[link] illustrates how these strategies are applied to an integrated-concept problem.

Example:
Recoil of a Dust Particle after Absorbing a Photon
The following topics are involved in this integrated concepts worked example:

Photons (quantum mechanics)
Linear Momentum

Topics

A 550-nm photon (visible light) is absorbed by a 1.00-μg particle of dust in outer space. (a) Find the momentum of such a photon. (b) What is the recoil velocity of the particle of dust, assuming it is initially at rest?

Strategy Step 1

To solve an *integrated-concept problem*, such as those following this example, we must first identify the physical principles involved and identify the chapters in which they are found. Part (a) of this example asks for the *momentum of a photon*, a topic of the present chapter. Part (b) considers *recoil following a collision*, a topic of Linear Momentum and Collisions.

Strategy Step 2

The following solutions to each part of the example illustrate how specific problem-solving strategies are applied. These involve identifying knowns and unknowns, checking to see if the answer is reasonable, and so on.

Solution for (a)

The momentum of a photon is related to its wavelength by the equation:

Equation:

$$p=\frac{h}{\lambda}.$$

Entering the known value for Planck's constant h and given the wavelength λ, we obtain

Equation:

$$\begin{aligned} p &= \frac{6.63\times 10^{-34}\text{ J}\cdot\text{s}}{550\times 10^{-9}\text{ m}} \\ &= 1.21\times 10^{-27}\text{ kg}\cdot\text{m/s}. \end{aligned}$$

Discussion for (a)

This momentum is small, as expected from discussions in the text and the fact that photons of visible light carry small amounts of energy and momentum compared with those carried by macroscopic objects.

Solution for (b)

Conservation of momentum in the absorption of this photon by a grain of dust can be analyzed using the equation:

Equation:

$$p_1+p_2=p\prime_1+p\prime_2(F_{\text{net}}=0).$$

The net external force is zero, since the dust is in outer space. Let 1 represent the photon and 2 the dust particle. Before the collision, the dust is at rest (relative to some observer); after the collision, there is no photon (it is absorbed). So conservation of momentum can be written

Equation:

$$p_1=p\prime_2=\text{mv},$$

where p_1 is the photon momentum before the collision and $p\prime_2$ is the dust momentum after the collision. The mass and recoil velocity of the dust are

m and v, respectively. Solving this for v, the requested quantity, yields
Equation:

$$v = \frac{p}{m},$$

where p is the photon momentum found in part (a). Entering known values (noting that a microgram is 10^{-9} kg) gives
Equation:

$$\begin{aligned} v &= \frac{1.21 \times 10^{-27}\ \text{kg}\cdot\text{m/s}}{1.00 \times 10^{-9}\ \text{kg}} \\ &= 1.21 \times 10^{-18}\ \text{m/s}. \end{aligned}$$

Discussion
The recoil velocity of the particle of dust is extremely small. As we have noted, however, there are immense numbers of photons in sunlight and other macroscopic sources. In time, collisions and absorption of many photons could cause a significant recoil of the dust, as observed in comet tails.

Section Summary

- The particle-wave duality refers to the fact that all particles—those with mass and those without mass—have wave characteristics.
- This is a further connection between mass and energy.

Conceptual Questions

Exercise:

Problem:
In what ways are matter and energy related that were not known before the development of relativity and quantum mechanics?

Problems & Exercises

Exercise:

Problem: Integrated Concepts

The 54.0-eV electron in [link] has a 0.167-nm wavelength. If such electrons are passed through a double slit and have their first maximum at an angle of 25.0^{o}, what is the slit separation d?

Solution:

0.395 nm

Exercise:

Problem: Integrated Concepts

An electron microscope produces electrons with a 2.00-pm wavelength. If these are passed through a 1.00-nm single slit, at what angle will the first diffraction minimum be found?

Exercise:

Problem: Integrated Concepts

A certain heat lamp emits 200 W of mostly IR radiation averaging 1500 nm in wavelength. (a) What is the average photon energy in joules? (b) How many of these photons are required to increase the temperature of a person’s shoulder by 2.0^{o}C, assuming the affected mass is 4.0 kg with a specific heat of $0.83\ \text{kcal/kg}{\cdot}^{\text{o}}\text{C}$. Also assume no other significant heat transfer. (c) How long does this take?

Solution:

(a) $1.3\times 10^{-19}\ \text{J}$

(b) 2.1×10^{23}

(c) 1.4×10^2 s

Exercise:

Problem: Integrated Concepts

On its high power setting, a microwave oven produces 900 W of 2560 MHz microwaves. (a) How many photons per second is this? (b) How many photons are required to increase the temperature of a 0.500-kg mass of pasta by $45.0^\circ\mathrm{C}$, assuming a specific heat of $0.900\ \mathrm{kcal/kg \cdot {}^\circ C}$? Neglect all other heat transfer. (c) How long must the microwave operator wait for their pasta to be ready?

Exercise:

Problem: Integrated Concepts

(a) Calculate the amount of microwave energy in joules needed to raise the temperature of 1.00 kg of soup from $20.0^\circ\mathrm{C}$ to $100^\circ\mathrm{C}$. (b) What is the total momentum of all the microwave photons it takes to do this? (c) Calculate the velocity of a 1.00-kg mass with the same momentum. (d) What is the kinetic energy of this mass?

Solution:

(a) $3.35 \times 10^5\ \mathrm{J}$

(b) $1.12 \times 10^{-3}\ \mathrm{kg \cdot m/s}$

(c) $1.12 \times 10^{-3}\ \mathrm{m/s}$

(d) $6.23 \times 10^{-7}\ \mathrm{J}$

Exercise:

Problem: Integrated Concepts

(a) What is γ for an electron emerging from the Stanford Linear Accelerator with a total energy of 50.0 GeV? (b) Find its momentum. (c) What is the electron's wavelength?

Exercise:

Problem: Integrated Concepts

(a) What is γ for a proton having an energy of 1.00 TeV, produced by the Fermilab accelerator? (b) Find its momentum. (c) What is the proton's wavelength?

Solution:

(a) 1.06×10^3

(b) $5.33 \times 10^{-16}\ \mathrm{kg \cdot m/s}$

(c) $1.24 \times 10^{-18}\ \mathrm{m}$

Exercise:

Problem: Integrated Concepts

An electron microscope passes 1.00-pm-wavelength electrons through a circular aperture $2.00\ \mu\mathrm{m}$ in diameter. What is the angle between two just-resolvable point sources for this microscope?

Exercise:

Problem: Integrated Concepts

(a) Calculate the velocity of electrons that form the same pattern as 450-nm light when passed through a double slit. (b) Calculate the kinetic energy of each and compare them. (c) Would either be easier to generate than the other? Explain.

Solution:

(a) $1.62 \times 10^{3}\ \mathrm{m/s}$

(b) 4.42×10^{-19} J for photon, 1.19×10^{-24} J for electron, photon energy is 3.71×10^{5} times greater

(c) The light is easier to make because 450-nm light is blue light and therefore easy to make. Creating electrons with $7.43\ \mu\text{eV}$ of energy would not be difficult, but would require a vacuum.

Exercise:

Problem: Integrated Concepts

(a) What is the separation between double slits that produces a second-order minimum at 45.0° for 650-nm light? (b) What slit separation is needed to produce the same pattern for 1.00-keV protons.

Solution:

(a) $2.30 \times 10^{-6}\ \mathrm{m}$

(b) $3.20 \times 10^{-12}\ \mathrm{m}$

Exercise:

Problem: Integrated Concepts

A laser with a power output of 2.00 mW at a wavelength of 400 nm is projected onto calcium metal. (a) How many electrons per second are ejected? (b) What power is carried away by the electrons, given that the binding energy is 2.71 eV? (c) Calculate the current of ejected electrons. (d) If the photoelectric material is electrically insulated and acts like a 2.00-pF capacitor, how long will current flow before the capacitor voltage stops it?

Exercise:

Problem: Integrated Concepts

One problem with x rays is that they are not sensed. Calculate the temperature increase of a researcher exposed in a few seconds to a nearly fatal accidental dose of x rays under the following conditions. The energy of the x-ray photons is 200 keV, and 4.00×10^{13} of them are absorbed per kilogram of tissue, the specific heat of which is $0.830 \text{ kcal/kg} \cdot \text{ºC}$. (Note that medical diagnostic x-ray machines *cannot* produce an intensity this great.)

Solution:

$3.69 \times 10^{-4} \text{ ºC}$

Exercise:

Problem: Integrated Concepts

A 1.00-fm photon has a wavelength short enough to detect some information about nuclei. (a) What is the photon momentum? (b) What is its energy in joules and MeV? (c) What is the (relativistic) velocity of an electron with the same momentum? (d) Calculate the electron's kinetic energy.

Exercise:

Problem: Integrated Concepts

The momentum of light is exactly reversed when reflected straight back from a mirror, assuming negligible recoil of the mirror. Thus the change in momentum is twice the photon momentum. Suppose light of intensity 1.00 kW/m^2 reflects from a mirror of area 2.00 m^2. (a) Calculate the energy reflected in 1.00 s. (b) What is the momentum imparted to the mirror? (c) Using the most general form of Newton's second law, what is the force on the mirror? (d) Does the assumption of no mirror recoil seem reasonable?

Solution:

(a) 2.00 kJ

(b) $1.33 \times 10^{-5}\ \mathrm{kg \cdot m/s}$

(c) $1.33 \times 10^{-5}\ \mathrm{N}$

(d) yes

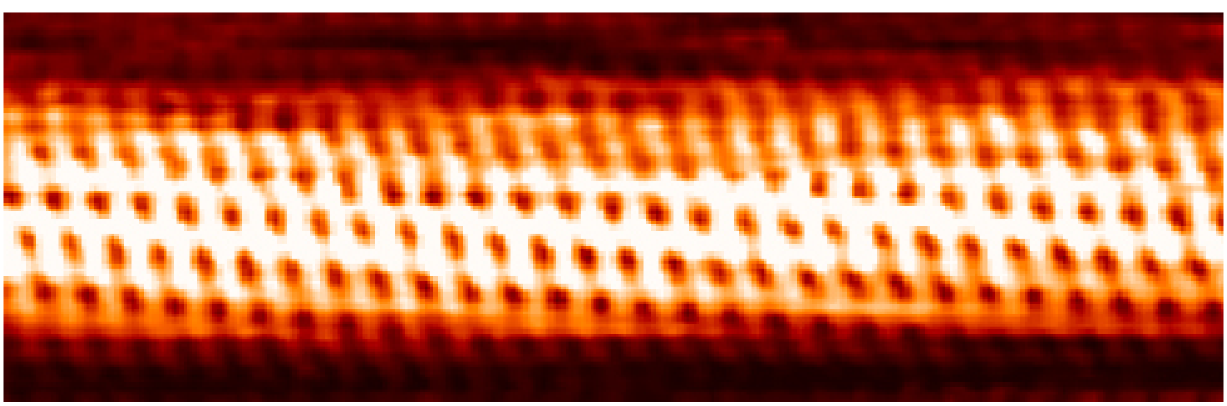
Exercise:

Problem: Integrated Concepts

Sunlight above the Earth's atmosphere has an intensity of $1.30\ \mathrm{kW/m^2}$. If this is reflected straight back from a mirror that has only a small recoil, the light's momentum is exactly reversed, giving the mirror twice the incident momentum. (a) Calculate the force per square meter of mirror. (b) Very low mass mirrors can be constructed in the near weightlessness of space, and attached to a spaceship to sail it. Once done, the average mass per square meter of the spaceship is 0.100 kg. Find the acceleration of the spaceship if all other forces are balanced. (c) How fast is it moving 24 hours later?

# Introduction to Atomic Physics
class="introduction"

Individual carbon atoms are visible in this image of a carbon nanotube made by a scanning tunneling electron microscope. (credit: Taner Yildirim, National Institute of Standards and Technology, via Wikimedia Commons)



From childhood on, we learn that atoms are a substructure of all things around us, from the air we breathe to the autumn leaves that blanket a forest trail. Invisible to the eye, the existence and properties of atoms are used to explain many phenomena—a theme found throughout this text. In this chapter, we discuss the discovery of atoms and their own substructures; we then apply quantum mechanics to the description of atoms, and their properties and interactions. Along the way, we will find, much like the scientists who made the original discoveries, that new concepts emerge with applications far beyond the boundaries of atomic physics.

Discovery of the Atom

- Describe the basic structure of the atom, the substructure of all matter.

How do we know that atoms are really there if we cannot see them with our eyes? A brief account of the progression from the proposal of atoms by the Greeks to the first direct evidence of their existence follows.

People have long speculated about the structure of matter and the existence of atoms. The earliest significant ideas to survive are due to the ancient Greeks in the fifth century BCE, especially those of the philosophers Leucippus and Democritus. (There is some evidence that philosophers in both India and China made similar speculations, at about the same time.) They considered the question of whether a substance can be divided without limit into ever smaller pieces. There are only a few possible answers to this question. One is that infinitesimally small subdivision is possible. Another is what Democritus in particular believed—that there is a smallest unit that cannot be further subdivided. Democritus called this the **atom**. We now know that atoms themselves can be subdivided, but their identity is destroyed in the process, so the Greeks were correct in a respect. The Greeks also felt that atoms were in constant motion, another correct notion.

The Greeks and others speculated about the properties of atoms, proposing that only a few types existed and that all matter was formed as various combinations of these types. The famous proposal that the basic elements were earth, air, fire, and water was brilliant, but incorrect. The Greeks had identified the most common examples of the four states of matter (solid, gas, plasma, and liquid), rather than the basic elements. More than 2000 years passed before observations could be made with equipment capable of revealing the true nature of atoms.

Over the centuries, discoveries were made regarding the properties of substances and their chemical reactions. Certain systematic features were recognized, but similarities between common and rare elements resulted in efforts to transmute them (lead into gold, in particular) for financial gain. Secrecy was endemic. Alchemists discovered and rediscovered many facts but did not make them broadly available. As the Middle Ages ended, alchemy gradually faded, and the science of chemistry arose. It was no

longer possible, nor considered desirable, to keep discoveries secret. Collective knowledge grew, and by the beginning of the 19th century, an important fact was well established—the masses of reactants in specific chemical reactions always have a particular mass ratio. This is very strong indirect evidence that there are basic units (atoms and molecules) that have these same mass ratios. The English chemist John Dalton (1766–1844) did much of this work, with significant contributions by the Italian physicist Amedeo Avogadro (1776–1856). It was Avogadro who developed the idea of a fixed number of atoms and molecules in a mole, and this special number is called Avogadro's number in his honor. The Austrian physicist Johann Josef Loschmidt was the first to measure the value of the constant in 1865 using the kinetic theory of gases.

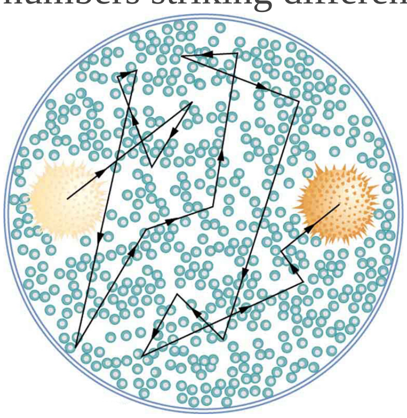
Note:
Patterns and Systematics
The recognition and appreciation of patterns has enabled us to make many discoveries. The periodic table of elements was proposed as an organized summary of the known elements long before all elements had been discovered, and it led to many other discoveries. We shall see in later chapters that patterns in the properties of subatomic particles led to the proposal of quarks as their underlying structure, an idea that is still bearing fruit.

Knowledge of the properties of elements and compounds grew, culminating in the mid-19th-century development of the periodic table of the elements by Dmitri Mendeleev (1834–1907), the great Russian chemist. Mendeleev proposed an ingenious array that highlighted the periodic nature of the properties of elements. Believing in the systematics of the periodic table, he also predicted the existence of then-unknown elements to complete it. Once these elements were discovered and determined to have properties predicted by Mendeleev, his periodic table became universally accepted.

Also during the 19th century, the kinetic theory of gases was developed. Kinetic theory is based on the existence of atoms and molecules in random

thermal motion and provides a microscopic explanation of the gas laws, heat transfer, and thermodynamics (see Introduction to Temperature, Kinetic Theory, and the Gas Laws and Introduction to Laws of Thermodynamics). Kinetic theory works so well that it is another strong indication of the existence of atoms. But it is still indirect evidence—individual atoms and molecules had not been observed. There were heated debates about the validity of kinetic theory until direct evidence of atoms was obtained.

The first truly direct evidence of atoms is credited to Robert Brown, a Scottish botanist. In 1827, he noticed that tiny pollen grains suspended in still water moved about in complex paths. This can be observed with a microscope for any small particles in a fluid. The motion is caused by the random thermal motions of fluid molecules colliding with particles in the fluid, and it is now called **Brownian motion**. (See [link].) Statistical fluctuations in the numbers of molecules striking the sides of a visible particle cause it to move first this way, then that. Although the molecules cannot be directly observed, their effects on the particle can be. By examining Brownian motion, the size of molecules can be calculated. The smaller and more numerous they are, the smaller the fluctuations in the numbers striking different sides.



The position of a pollen grain in water, measured every few seconds under a microscope,

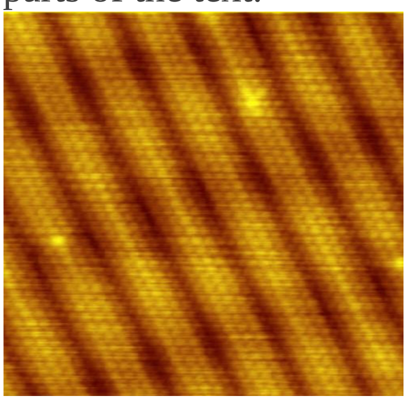
exhibits Brownian motion. Brownian motion is due to fluctuations in the number of atoms and molecules colliding with a small mass, causing it to move about in complex paths. This is nearly direct evidence for the existence of atoms, providing a satisfactory alternative explanation cannot be found.

It was Albert Einstein who, starting in his epochal year of 1905, published several papers that explained precisely how Brownian motion could be used to measure the size of atoms and molecules. (In 1905 Einstein created special relativity, proposed photons as quanta of EM radiation, and produced a theory of Brownian motion that allowed the size of atoms to be determined. All of this was done in his spare time, since he worked days as a patent examiner. Any one of these very basic works could have been the crowning achievement of an entire career—yet Einstein did even more in later years.) Their sizes were only approximately known to be 10^{-10} m, based on a comparison of latent heat of vaporization and surface tension made in about 1805 by Thomas Young of double-slit fame and the famous astronomer and mathematician Simon Laplace.

Using Einstein's ideas, the French physicist Jean-Baptiste Perrin (1870–1942) carefully observed Brownian motion; not only did he confirm Einstein's theory, he also produced accurate sizes for atoms and molecules.

Since molecular weights and densities of materials were well established, knowing atomic and molecular sizes allowed a precise value for Avogadro's number to be obtained. (If we know how big an atom is, we know how many fit into a certain volume.) Perrin also used these ideas to explain atomic and molecular agitation effects in sedimentation, and he received the 1926 Nobel Prize for his achievements. Most scientists were already convinced of the existence of atoms, but the accurate observation and analysis of Brownian motion was conclusive—it was the first truly direct evidence.

A huge array of direct and indirect evidence for the existence of atoms now exists. For example, it has become possible to accelerate ions (much as electrons are accelerated in cathode-ray tubes) and to detect them individually as well as measure their masses (see More Applications of Magnetism for a discussion of mass spectrometers). Other devices that observe individual atoms, such as the scanning tunneling electron microscope, will be discussed elsewhere. (See [link].) All of our understanding of the properties of matter is based on and consistent with the atom. The atom's substructures, such as electron shells and the nucleus, are both interesting and important. The nucleus in turn has a substructure, as do the particles of which it is composed. These topics, and the question of whether there is a smallest basic structure to matter, will be explored in later parts of the text.



Individual atoms can be detected with devices such as the scanning tunneling electron

microscope that produced this image of individual gold atoms on a graphite substrate. (credit: Erwin Rossen, Eindhoven University of Technology, via Wikimedia Commons)

Section Summary

- Atoms are the smallest unit of elements; atoms combine to form molecules, the smallest unit of compounds.
- The first direct observation of atoms was in Brownian motion.
- Analysis of Brownian motion gave accurate sizes for atoms (10^{-10} m on average) and a precise value for Avogadro's number.

Conceptual Questions

Exercise:

Problem:

Name three different types of evidence for the existence of atoms.

Exercise:

Problem:

Explain why patterns observed in the periodic table of the elements are evidence for the existence of atoms, and why Brownian motion is a more direct type of evidence for their existence.

Exercise:

Problem: If atoms exist, why can’t we see them with visible light?

Problems & Exercises

Exercise:

Problem:

Using the given charge-to-mass ratios for electrons and protons, and knowing the magnitudes of their charges are equal, what is the ratio of the proton’s mass to the electron’s? (Note that since the charge-to-mass ratios are given to only three-digit accuracy, your answer may differ from the accepted ratio in the fourth digit.)

Solution:

1.84×10^3

Exercise:

Problem:

(a) Calculate the mass of a proton using the charge-to-mass ratio given for it in this chapter and its known charge. (b) How does your result compare with the proton mass given in this chapter?

Exercise:

Problem:

If someone wanted to build a scale model of the atom with a nucleus 1.00 m in diameter, how far away would the nearest electron need to be?

Solution:

50 km

Glossary

atom
: basic unit of matter, which consists of a central, positively charged nucleus surrounded by negatively charged electrons

Brownian motion
: the continuous random movement of particles of matter suspended in a liquid or gas

Discovery of the Parts of the Atom: Electrons and Nuclei

- Describe how electrons were discovered.
- Explain the Millikan oil drop experiment.
- Describe Rutherford’s gold foil experiment.
- Describe Rutherford’s planetary model of the atom.

Just as atoms are a substructure of matter, electrons and nuclei are substructures of the atom. The experiments that were used to discover electrons and nuclei reveal some of the basic properties of atoms and can be readily understood using ideas such as electrostatic and magnetic force, already covered in previous chapters.

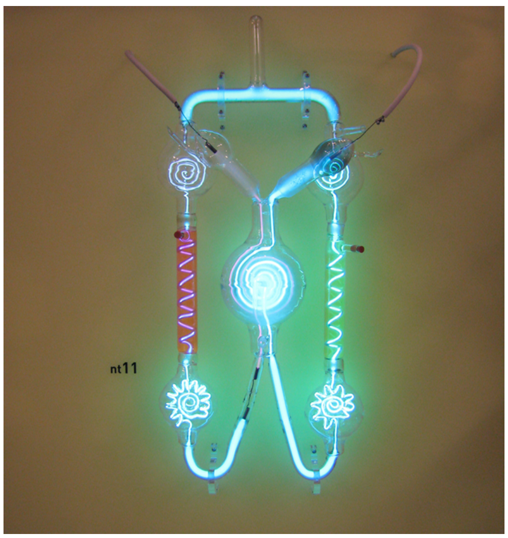
Note:
Charges and Electromagnetic Forces
In previous discussions, we have noted that positive charge is associated with nuclei and negative charge with electrons. We have also covered many aspects of the electric and magnetic forces that affect charges. We will now explore the discovery of the electron and nucleus as substructures of the atom and examine their contributions to the properties of atoms.

The Electron

Gas discharge tubes, such as that shown in [link], consist of an evacuated glass tube containing two metal electrodes and a rarefied gas. When a high voltage is applied to the electrodes, the gas glows. These tubes were the precursors to today’s neon lights. They were first studied seriously by Heinrich Geissler, a German inventor and glassblower, starting in the 1860s. The English scientist William Crookes, among others, continued to study what for some time were called Crookes tubes, wherein electrons are freed from atoms and molecules in the rarefied gas inside the tube and are accelerated from the cathode (negative) to the anode (positive) by the high potential. These “*cathode rays*” collide with the gas atoms and molecules and excite them, resulting in the emission of electromagnetic (EM)

radiation that makes the electrons’ path visible as a ray that spreads and fades as it moves away from the cathode.

Gas discharge tubes today are most commonly called **cathode-ray tubes**, because the rays originate at the cathode. Crookes showed that the electrons carry momentum (they can make a small paddle wheel rotate). He also found that their normally straight path is bent by a magnet in the direction expected for a negative charge moving away from the cathode. These were the first direct indications of electrons and their charge.



A gas discharge tube glows when a high voltage is applied to it. Electrons emitted from the cathode are accelerated toward the anode; they excite atoms and molecules in the gas, which glow in response. Once called Geissler tubes and later Crookes tubes, they are now known as cathode-ray

tubes (CRTs) and are found in older TVs, computer screens, and x-ray machines. When a magnetic field is applied, the beam bends in the direction expected for negative charge. (credit: Paul Downey, Flickr)

The English physicist J. J. Thomson (1856–1940) improved and expanded the scope of experiments with gas discharge tubes. (See [link] and [link].) He verified the negative charge of the cathode rays with both magnetic and electric fields. Additionally, he collected the rays in a metal cup and found an excess of negative charge. Thomson was also able to measure the ratio of the charge of the electron to its mass, q_e/m_e—an important step to finding the actual values of both q_e and m_e. [link] shows a cathode-ray tube, which produces a narrow beam of electrons that passes through charging plates connected to a high-voltage power supply. An electric field $\mathbf{E}$ is produced between the charging plates, and the cathode-ray tube is placed between the poles of a magnet so that the electric field $\mathbf{E}$ is perpendicular to the magnetic field $\mathbf{B}$ of the magnet. These fields, being perpendicular to each other, produce opposing forces on the electrons. As discussed for mass spectrometers in More Applications of Magnetism, if the net force due to the fields vanishes, then the velocity of the charged particle is $v = E/B$. In this manner, Thomson determined the velocity of the electrons and then moved the beam up and down by adjusting the electric field.

J. J. Thomson (credit: www.firstworldwar.com , via Wikimedia Commons)

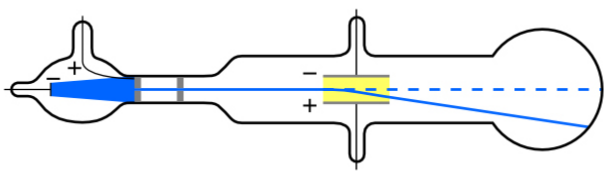
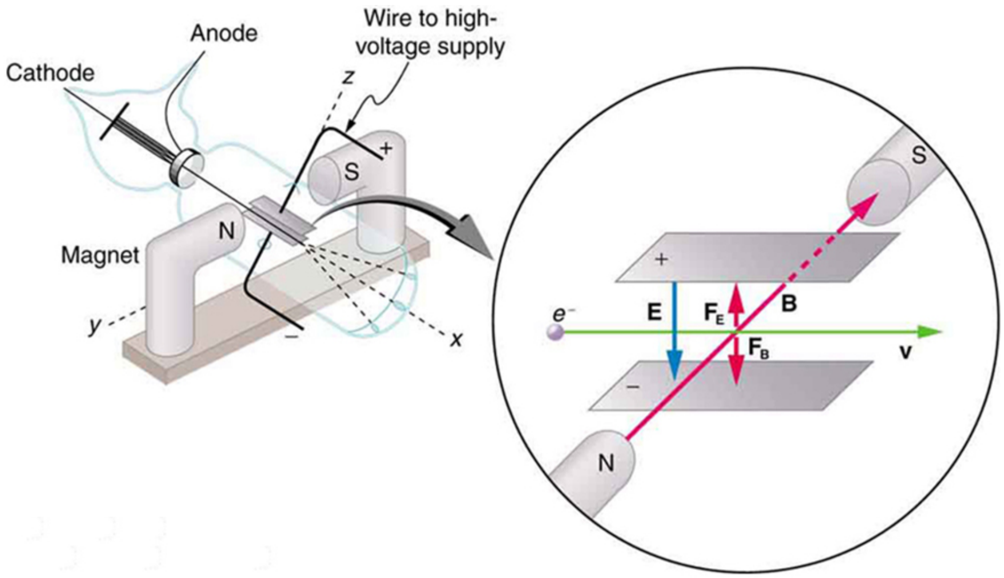


Diagram of Thomson’s CRT. (credit: Kurzon, Wikimedia Commons)



This schematic shows the electron beam in a CRT passing through crossed electric and magnetic fields and causing phosphor to glow when striking the end of the tube.

To see how the amount of deflection is used to calculate q_e/m_e, note that the deflection is proportional to the electric force on the electron:

Equation:

$$F = q_e E.$$

But the vertical deflection is also related to the electron's mass, since the electron's acceleration is

Equation:

$$a = \frac{F}{m_e}.$$

The value of F is not known, since q_e was not yet known. Substituting the expression for electric force into the expression for acceleration yields

Equation:

$$a=\frac{F}{m_e}=\frac{q_eE}{m_e}.$$

Gathering terms, we have
Equation:

$$\frac{q_e}{m_e}=\frac{a}{E}.$$

The deflection is analyzed to get a, and E is determined from the applied voltage and distance between the plates; thus, $\frac{q_e}{m_e}$ can be determined. With the velocity known, another measurement of $\frac{q_e}{m_e}$ can be obtained by bending the beam of electrons with the magnetic field. Since $F_{\text{mag}}=q_e\text{vB}=m_ea$, we have $q_e/m_e=a/\text{vB}$. Consistent results are obtained using magnetic deflection.

What is so important about q_e/m_e, the ratio of the electron's charge to its mass? The value obtained is
Equation:

$$\frac{q_e}{m_e}=-1.76\times10^{11}\ \text{C/kg (electron)}.$$

This is a huge number, as Thomson realized, and it implies that the electron has a very small mass. It was known from electroplating that about $10^8\ \text{C/kg}$ is needed to plate a material, a factor of about 1000 less than the charge per kilogram of electrons. Thomson went on to do the same experiment for positively charged hydrogen ions (now known to be bare protons) and found a charge per kilogram about 1000 times smaller than that for the electron, implying that the proton is about 1000 times more massive than the electron. Today, we know more precisely that
Equation:

$$\frac{q_p}{m_p} = 9.58 \times 10^7 \text{ C/kg (proton)},$$

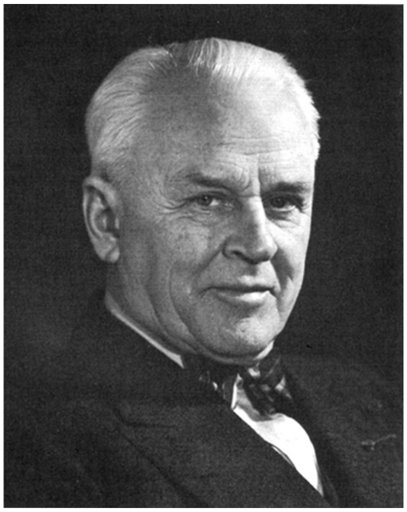
where q_p is the charge of the proton and m_p is its mass. This ratio (to four significant figures) is 1836 times less charge per kilogram than for the electron. Since the charges of electrons and protons are equal in magnitude, this implies $m_p = 1836 m_e$.

Thomson performed a variety of experiments using differing gases in discharge tubes and employing other methods, such as the photoelectric effect, for freeing electrons from atoms. He always found the same properties for the electron, proving it to be an independent particle. For his work, the important pieces of which he began to publish in 1897, Thomson was awarded the 1906 Nobel Prize in Physics. In retrospect, it is difficult to appreciate how astonishing it was to find that the atom has a substructure. Thomson himself said, “It was only when I was convinced that the experiment left no escape from it that I published my belief in the existence of bodies smaller than atoms.”

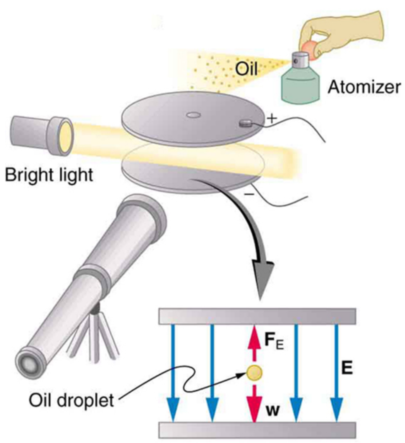
Thomson attempted to measure the charge of individual electrons, but his method could determine its charge only to the order of magnitude expected.

Since Faraday’s experiments with electroplating in the 1830s, it had been known that about 100,000 C per mole was needed to plate singly ionized ions. Dividing this by the number of ions per mole (that is, by Avogadro’s number), which was approximately known, the charge per ion was calculated to be about 1.6×10^{-19} C, close to the actual value.

An American physicist, Robert Millikan (1868–1953) (see [link]), decided to improve upon Thomson’s experiment for measuring q_e and was eventually forced to try another approach, which is now a classic experiment performed by students. The Millikan oil drop experiment is shown in [link].



Robert Millikan (credit: Unknown Author, via Wikimedia Commons)



The Millikan oil drop experiment produced the first accurate direct measurement of the

charge on electrons, one of the most fundamental constants in nature. Fine drops of oil become charged when sprayed. Their movement is observed between metal plates with a potential applied to oppose the gravitational force. The balance of gravitational and electric forces allows the calculation of the charge on a drop. The charge is found to be quantized in units of -1.6×10^{-19} C, thus determining directly the charge of the excess and missing electrons on the oil drops.

In the Millikan oil drop experiment, fine drops of oil are sprayed from an atomizer. Some of these are charged by the process and can then be suspended between metal plates by a voltage between the plates. In this situation, the weight of the drop is balanced by the electric force:

Equation:

$$m_{\text{drop}}g = q_e E$$

The electric field is produced by the applied voltage, hence, $E = V/d$, and V is adjusted to just balance the drop's weight. The drops can be seen as points of reflected light using a microscope, but they are too small to directly measure their size and mass. The mass of the drop is determined by observing how fast it falls when the voltage is turned off. Since air resistance is very significant for these submicroscopic drops, the more massive drops fall faster than the less massive, and sophisticated sedimentation calculations can reveal their mass. Oil is used rather than water, because it does not readily evaporate, and so mass is nearly constant. Once the mass of the drop is known, the charge of the electron is given by rearranging the previous equation:

Equation:

$$q = \frac{m_{\text{drop}}g}{E} = \frac{m_{\text{drop}}\text{gd}}{V},$$

where d is the separation of the plates and V is the voltage that holds the drop motionless. (The same drop can be observed for several hours to see that it really is motionless.) By 1913 Millikan had measured the charge of the electron q_e to an accuracy of 1%, and he improved this by a factor of 10 within a few years to a value of -1.60×10^{-19} C. He also observed that all charges were multiples of the basic electron charge and that sudden changes could occur in which electrons were added or removed from the drops. For this very fundamental direct measurement of q_e and for his studies of the photoelectric effect, Millikan was awarded the 1923 Nobel Prize in Physics.

With the charge of the electron known and the charge-to-mass ratio known, the electron's mass can be calculated. It is

Equation:

$$m = \frac{q_e}{\left(\frac{q_e}{m_e}\right)}.$$

Substituting known values yields
Equation:

$$m_e = \frac{-1.60 \times 10^{-19}\ \text{C}}{-1.76 \times 10^{11}\ \text{C/kg}}$$

or
Equation:

$$m_e = 9.11 \times 10^{-31}\ \text{kg}\ \ (\text{electron's mass}),$$

where the round-off errors have been corrected. The mass of the electron has been verified in many subsequent experiments and is now known to an accuracy of better than one part in one million. It is an incredibly small mass and remains the smallest known mass of any particle that has mass. (Some particles, such as photons, are massless and cannot be brought to rest, but travel at the speed of light.) A similar calculation gives the masses of other particles, including the proton. To three digits, the mass of the proton is now known to be
Equation:

$$m_p = 1.67 \times 10^{-27}\ \text{kg}\ \ (\text{proton's mass}),$$

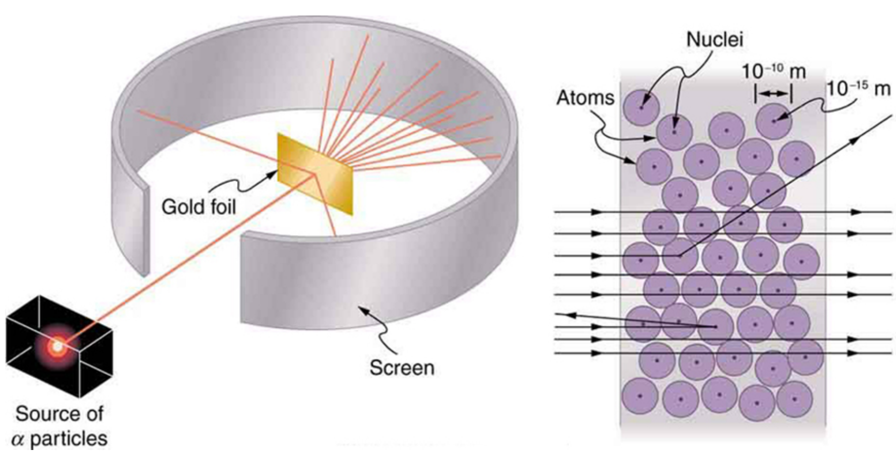
which is nearly identical to the mass of a hydrogen atom. What Thomson and Millikan had done was to prove the existence of one substructure of atoms, the electron, and further to show that it had only a tiny fraction of the mass of an atom. The nucleus of an atom contains most of its mass, and the nature of the nucleus was completely unanticipated.

Another important characteristic of quantum mechanics was also beginning to emerge. All electrons are identical to one another. The charge and mass of electrons are not average values; rather, they are unique values that all electrons have. This is true of other fundamental entities at the submicroscopic level. All protons are identical to one another, and so on.

The Nucleus

Here, we examine the first direct evidence of the size and mass of the nucleus. In later chapters, we will examine many other aspects of nuclear physics, but the basic information on nuclear size and mass is so important to understanding the atom that we consider it here.

Nuclear radioactivity was discovered in 1896, and it was soon the subject of intense study by a number of the best scientists in the world. Among them was New Zealander Lord Ernest Rutherford, who made numerous fundamental discoveries and earned the title of “father of nuclear physics.” Born in Nelson, Rutherford did his postgraduate studies at the Cavendish Laboratories in England before taking up a position at McGill University in Canada where he did the work that earned him a Nobel Prize in Chemistry in 1908. In the area of atomic and nuclear physics, there is much overlap between chemistry and physics, with physics providing the fundamental enabling theories. He returned to England in later years and had six future Nobel Prize winners as students. Rutherford used nuclear radiation to directly examine the size and mass of the atomic nucleus. The experiment he devised is shown in [link]. A radioactive source that emits alpha radiation was placed in a lead container with a hole in one side to produce a beam of alpha particles, which are a type of ionizing radiation ejected by the nuclei of a radioactive source. A thin gold foil was placed in the beam, and the scattering of the alpha particles was observed by the glow they caused when they struck a phosphor screen.



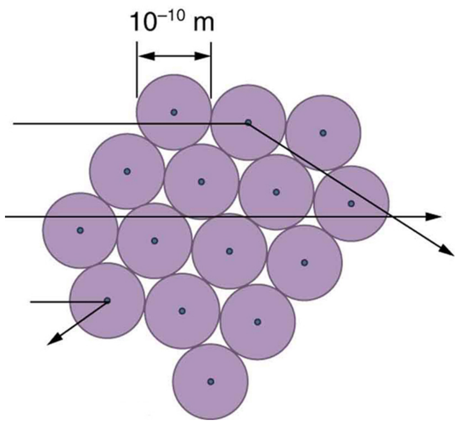
Rutherford's experiment gave direct evidence for the size and mass of the nucleus by scattering alpha particles from a thin gold foil. Alpha particles with energies of about $5\ \mathrm{MeV}$ are emitted from a radioactive source (which is a small metal container in which a specific amount of a radioactive material is sealed), are collimated into a beam, and fall upon the foil. The number of particles that penetrate the foil or scatter to various angles indicates that gold nuclei are very small and contain nearly all of the gold atom's mass. This is particularly indicated by the alpha particles that scatter to very large angles, much like a soccer ball bouncing off a goalie's head.

Alpha particles were known to be the doubly charged positive nuclei of helium atoms that had kinetic energies on the order of $5\ \mathrm{MeV}$ when emitted in nuclear decay, which is the disintegration of the nucleus of an unstable nuclide by the spontaneous emission of charged particles. These particles interact with matter mostly via the Coulomb force, and the manner in which they scatter from nuclei can reveal nuclear size and mass. This is analogous to observing how a bowling ball is scattered by an object you cannot see directly. Because the alpha particle's energy is so large compared with the typical energies associated with atoms (MeV versus eV), you would expect the alpha particles to simply crash through a thin foil much like a supersonic bowling ball would crash through a few dozen rows of bowling pins. Thomson had envisioned the atom to be a small sphere in which equal amounts of positive and negative charge were distributed evenly. The incident massive alpha particles would suffer only small deflections in such a model. Instead, Rutherford and his collaborators found that alpha particles occasionally were scattered to large angles, some even back in the direction from which they came! Detailed analysis using conservation of momentum and energy—particularly of the small number that came straight back—implied that gold nuclei are very small compared with the size of a gold atom, contain almost all of the atom's mass, and are tightly bound. Since

the gold nucleus is several times more massive than the alpha particle, a head-on collision would scatter the alpha particle straight back toward the source. In addition, the smaller the nucleus, the fewer alpha particles that would hit one head on.

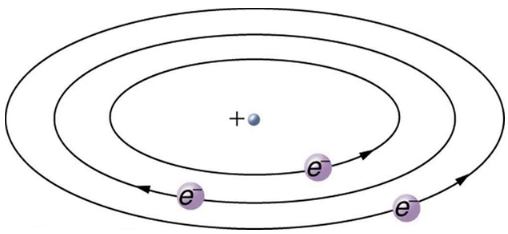
Although the results of the experiment were published by his colleagues in 1909, it took Rutherford two years to convince himself of their meaning. Like Thomson before him, Rutherford was reluctant to accept such radical results. Nature on a small scale is so unlike our classical world that even those at the forefront of discovery are sometimes surprised. Rutherford later wrote: “It was almost as incredible as if you fired a 15-inch shell at a piece of tissue paper and it came back and hit you. On consideration, I realized that this scattering backwards ... [meant] ... the greatest part of the mass of the atom was concentrated in a tiny nucleus.” In 1911, Rutherford published his analysis together with a proposed model of the atom. The size of the nucleus was determined to be about $10^{-15}\ \text{m}$, or 100,000 times smaller than the atom. This implies a huge density, on the order of $10^{15}\ \text{g/cm}^3$, vastly unlike any macroscopic matter. Also implied is the existence of previously unknown nuclear forces to counteract the huge repulsive Coulomb forces among the positive charges in the nucleus. Huge forces would also be consistent with the large energies emitted in nuclear radiation.

The small size of the nucleus also implies that the atom is mostly empty inside. In fact, in Rutherford’s experiment, most alphas went straight through the gold foil with very little scattering, since electrons have such small masses and since the atom was mostly empty with nothing for the alpha to hit. There were already hints of this at the time Rutherford performed his experiments, since energetic electrons had been observed to penetrate thin foils more easily than expected. [link] shows a schematic of the atoms in a thin foil with circles representing the size of the atoms (about $10^{-10}\ \text{m}$) and dots representing the nuclei. (The dots are not to scale—if they were, you would need a microscope to see them.) Most alpha particles miss the small nuclei and are only slightly scattered by electrons. Occasionally, (about once in 8000 times in Rutherford’s experiment), an alpha hits a nucleus head-on and is scattered straight backward.



An expanded view of the atoms in the gold foil in Rutherford’s experiment. Circles represent the atoms (about 10^{-10} m in diameter), while the dots represent the nuclei (about 10^{-15} m in diameter). To be visible, the dots are much larger than scale. Most alpha particles crash through but are relatively unaffected because of their high energy and the electron’s small mass. Some, however, head straight toward a nucleus and are scattered straight back. A detailed analysis gives the size and mass of the nucleus.

Based on the size and mass of the nucleus revealed by his experiment, as well as the mass of electrons, Rutherford proposed the **planetary model of the atom**. The planetary model of the atom pictures low-mass electrons orbiting a large-mass nucleus. The sizes of the electron orbits are large compared with the size of the nucleus, with mostly vacuum inside the atom. This picture is analogous to how low-mass planets in our solar system orbit the large-mass Sun at distances large compared with the size of the sun. In the atom, the attractive Coulomb force is analogous to gravitation in the planetary system. (See [link].) Note that a model or mental picture is needed to explain experimental results, since the atom is too small to be directly observed with visible light.



Rutherford's planetary model of the atom incorporates the characteristics of the nucleus, electrons, and the size of the atom. This model was the first to recognize the structure of atoms, in which low-mass electrons orbit a very small, massive nucleus in orbits much larger than the nucleus. The atom is mostly empty and is analogous to our planetary system.

Rutherford’s planetary model of the atom was crucial to understanding the characteristics of atoms, and their interactions and energies, as we shall see in the next few sections. Also, it was an indication of how different nature is from the familiar classical world on the small, quantum mechanical scale. The discovery of a substructure to all matter in the form of atoms and molecules was now being taken a step further to reveal a substructure of atoms that was simpler than the 92 elements then known. We have continued to search for deeper substructures, such as those inside the nucleus, with some success. In later chapters, we will follow this quest in the discussion of quarks and other elementary particles, and we will look at the direction the search seems now to be heading.

Note:
PhET Explorations: Rutherford Scattering
How did Rutherford figure out the structure of the atom without being able to see it? Simulate the famous experiment in which he disproved the Plum Pudding model of the atom by observing alpha particles bouncing off atoms and determining that they must have a small core.

https://phet.colorado.edu/sims/html/rutherford-scattering/latest/rutherford-scattering_en.html

Section Summary

- Atoms are composed of negatively charged electrons, first proved to exist in cathode-ray-tube experiments, and a positively charged nucleus.
- All electrons are identical and have a charge-to-mass ratio of
  Equation:

$$\frac{q_e}{m_e} = -1.76 \times 10^{11}\ \text{C/kg}.$$

- The positive charge in the nuclei is carried by particles called protons, which have a charge-to-mass ratio of
  Equation:

$$\frac{q_p}{m_p} = 9.57 \times 10^7 \text{ C/kg.}$$

- Mass of electron,
  Equation:

$$m_e = 9.11 \times 10^{-31} \text{ kg.}$$

- Mass of proton,
  Equation:

$$m_p = 1.67 \times 10^{-27} \text{ kg.}$$

- The planetary model of the atom pictures electrons orbiting the nucleus in the same way that planets orbit the sun.

Conceptual Questions

Exercise:

Problem:

What two pieces of evidence allowed the first calculation of m_e, the mass of the electron?

(a) The ratios q_e/m_e and q_p/m_p.

(b) The values of q_e and E_B.

(c) The ratio q_e/m_e and q_e.

Justify your response.

Exercise:

Problem:

How do the allowed orbits for electrons in atoms differ from the allowed orbits for planets around the sun? Explain how the correspondence principle applies here.

Problem Exercises

Exercise:

Problem:

Rutherford found the size of the nucleus to be about $10^{-15}\ \text{m}$. This implied a huge density. What would this density be for gold?

Solution:

$6 \times 10^{20}\ \text{kg/m}^3$

Exercise:

Problem:

In Millikan's oil-drop experiment, one looks at a small oil drop held motionless between two plates. Take the voltage between the plates to be 2033 V, and the plate separation to be 2.00 cm. The oil drop (of density $0.81\ \text{g/cm}^3$) has a diameter of $4.0 \times 10^{-6}\ \text{m}$. Find the charge on the drop, in terms of electron units.

Exercise:

Problem:

(a) An aspiring physicist wants to build a scale model of a hydrogen atom for her science fair project. If the atom is 1.00 m in diameter, how big should she try to make the nucleus?

(b) How easy will this be to do?

Solution:

(a) 10.0 μm

(b) It isn't hard to make one of approximately this size. It would be harder to make it exactly 10.0 μm.

Glossary

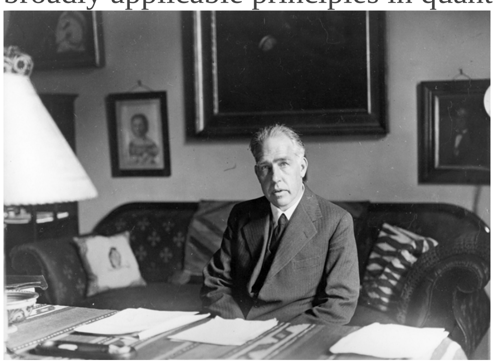
cathode-ray tube
: a vacuum tube containing a source of electrons and a screen to view images

planetary model of the atom
: the most familiar model or illustration of the structure of the atom

Bohr’s Theory of the Hydrogen Atom

- Describe the mysteries of atomic spectra.
- Explain Bohr’s theory of the hydrogen atom.
- Explain Bohr’s planetary model of the atom.
- Illustrate energy state using the energy-level diagram.
- Describe the triumphs and limits of Bohr’s theory.

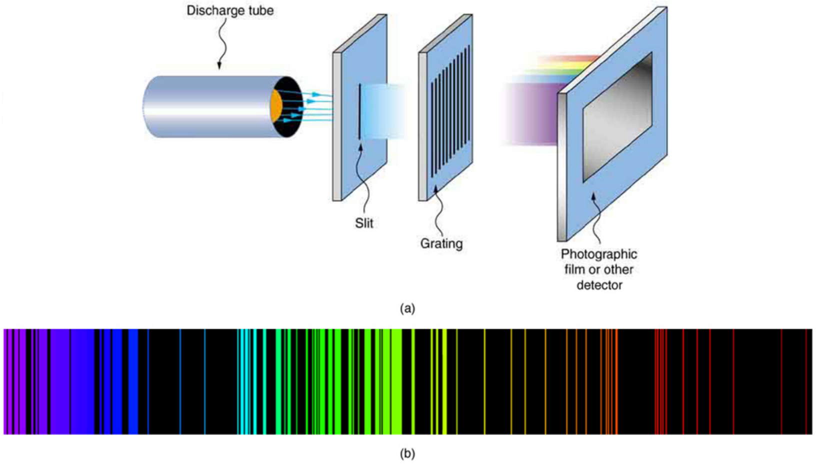
The great Danish physicist Niels Bohr (1885–1962) made immediate use of Rutherford’s planetary model of the atom. ([link]). Bohr became convinced of its validity and spent part of 1912 at Rutherford’s laboratory. In 1913, after returning to Copenhagen, he began publishing his theory of the simplest atom, hydrogen, based on the planetary model of the atom. For decades, many questions had been asked about atomic characteristics. From their sizes to their spectra, much was known about atoms, but little had been explained in terms of the laws of physics. Bohr’s theory explained the atomic spectrum of hydrogen and established new and broadly applicable principles in quantum mechanics.



Niels Bohr, Danish physicist, used the planetary model of the atom to explain the atomic spectrum and size of the hydrogen atom. His many contributions to the development of atomic physics and quantum mechanics, his personal influence on many students and colleagues, and his personal integrity, especially in the face of Nazi oppression, earned him a prominent place in history. (credit: Unknown Author, via Wikimedia Commons)

Mysteries of Atomic Spectra

As noted in Quantization of Energy , the energies of some small systems are quantized. Atomic and molecular emission and absorption spectra have been known for over a century to be discrete (or quantized). (See [link].) Maxwell and others had realized that there must be a connection between the spectrum of an atom and its structure, something like the resonant frequencies of musical instruments. But, in spite of years of efforts by many great minds, no one had a workable theory. (It was a running joke that any theory of atomic and molecular spectra could be destroyed by throwing a book of data at it, so complex were the spectra.) Following Einstein's proposal of photons with quantized energies directly proportional to their wavelengths, it became even more evident that electrons in atoms can exist only in discrete orbits.



Part (a) shows, from left to right, a discharge tube, slit, and diffraction grating producing a line spectrum. Part (b) shows the emission line spectrum for iron. The discrete lines imply quantized energy states for the atoms that produce them. The line spectrum for each element is unique, providing a powerful and much used analytical tool, and many line spectra were well known for many years before they could be explained with physics. (credit for (b): Yttrium91, Wikimedia Commons)

In some cases, it had been possible to devise formulas that described the emission spectra. As you might expect, the simplest atom—hydrogen, with its single electron—has a relatively simple spectrum. The hydrogen spectrum had been observed in the infrared (IR), visible, and ultraviolet (UV), and several series of spectral lines had been observed. (See [link].) These series are named after early researchers who studied them in particular depth.

The observed **hydrogen-spectrum wavelengths** can be calculated using the following formula:

Equation:

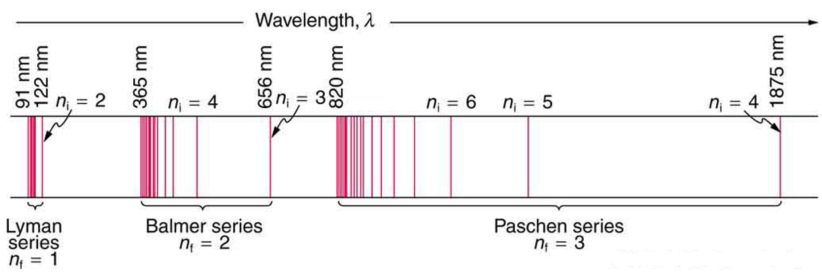
$$\frac{1}{\lambda}=R\left(\frac{1}{n_{\mathrm{f}}^{2}}-\frac{1}{n_{\mathrm{i}}^{2}}\right),$$

where λ is the wavelength of the emitted EM radiation and R is the **Rydberg constant**, determined by the experiment to be

Equation:

$$R=1.097\times10^{7}/\mathrm{m}\ (\mathrm{or}\ \mathrm{m}^{-1}).$$

The constant n_{f} is a positive integer associated with a specific series. For the Lyman series, $n_{\mathrm{f}}=1$; for the Balmer series, $n_{\mathrm{f}}=2$; for the Paschen series, $n_{\mathrm{f}}=3$; and so on. The Lyman series is entirely in the UV, while part of the Balmer series is visible with the remainder UV. The Paschen series and all the rest are entirely IR. There are apparently an unlimited number of series, although they lie progressively farther into the infrared and become difficult to observe as n_{f} increases. The constant n_{i} is a positive integer, but it must be greater than n_{f}. Thus, for the Balmer series, $n_{\mathrm{f}}=2$ and $n_{\mathrm{i}}=3, 4, 5, 6, ...$ Note that n_{i} can approach infinity. While the formula in the wavelengths equation was just a recipe designed to fit data and was not based on physical principles, it did imply a deeper meaning. Balmer first devised the formula for his series alone, and it was later found to describe all the other series by using different values of n_{f}. Bohr was the first to comprehend the deeper meaning. Again, we see the interplay between experiment and theory in physics. Experimentally, the spectra were well established, an equation was found to fit the experimental data, but the theoretical foundation was missing.



A schematic of the hydrogen spectrum shows several series named for those who contributed most to their determination. Part of the Balmer series is in the visible spectrum, while the Lyman series is entirely in the UV, and the Paschen series and others are in the IR. Values of n_{f} and n_{i} are shown for some of the lines.

Example:
Calculating Wave Interference of a Hydrogen Line

What is the distance between the slits of a grating that produces a first-order maximum for the second Balmer line at an angle of 15º?

Strategy and Concept

For an Integrated Concept problem, we must first identify the physical principles involved. In this example, we need to know (a) the wavelength of light as well as (b) conditions for an interference maximum for the pattern from a double slit. Part (a) deals with a topic of the present chapter, while part (b) considers the wave interference material of Wave Optics.

Solution for (a)

Hydrogen spectrum wavelength. The Balmer series requires that $n_{\mathrm{f}} = 2$. The first line in the series is taken to be for $n_{\mathrm{i}} = 3$, and so the second would have $n_{\mathrm{i}} = 4$.

The calculation is a straightforward application of the wavelength equation. Entering the determined values for n_{f} and n_{i} yields

Equation:

$$\begin{aligned} \frac{1}{\lambda} &= R\left(\frac{1}{n_{\mathrm{f}}^2} - \frac{1}{n_{\mathrm{i}}^2}\right) \\ &= \left(1.097 \times 10^7\ \mathrm{m}^{-1}\right)\left(\frac{1}{2^2} - \frac{1}{4^2}\right) \\ &= 2.057 \times 10^6\ \mathrm{m}^{-1}. \end{aligned}$$

Inverting to find λ gives

Equation:

$$\begin{aligned} \lambda &= \frac{1}{2.057 \times 10^6\ \mathrm{m}^{-1}} = 486 \times 10^{-9}\ \mathrm{m} \\ &= 486\ \mathrm{nm}. \end{aligned}$$

Discussion for (a)

This is indeed the experimentally observed wavelength, corresponding to the second (blue-green) line in the Balmer series. More impressive is the fact that the same simple recipe predicts *all* of the hydrogen spectrum lines, including new ones observed in subsequent experiments. What is nature telling us?

Solution for (b)

Double-slit interference (Wave Optics). To obtain constructive interference for a double slit, the path length difference from two slits must be an integral multiple of the wavelength. This condition was expressed by the equation

Equation:

$$d \sin \theta = m\lambda,$$

where d is the distance between slits and θ is the angle from the original direction of the beam. The number m is the order of the interference; $m = 1$ in this example. Solving for d and entering known values yields

Equation:

$$d = \frac{(1)(486\text{ nm})}{\sin 15^{\circ}} = 1.88 \times 10^{-6}\text{ m}.$$

Discussion for (b)

This number is similar to those used in the interference examples of Introduction to Quantum Physics (and is close to the spacing between slits in commonly used diffraction glasses).

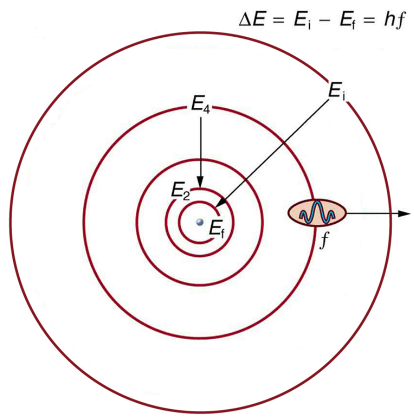
Bohr's Solution for Hydrogen

Bohr was able to derive the formula for the hydrogen spectrum using basic physics, the planetary model of the atom, and some very important new proposals. His first proposal is that only certain orbits are allowed: we say that *the orbits of electrons in atoms are quantized*. Each orbit has a different energy, and electrons can move to a higher orbit by absorbing energy and drop to a lower orbit by emitting energy. If the orbits are quantized, the amount of energy absorbed or emitted is also quantized, producing discrete spectra. Photon absorption and emission are among the primary methods of transferring energy into and out of atoms. The energies of the photons are quantized, and their energy is explained as being equal to the change in energy of the electron when it moves from one orbit to another. In equation form, this is

Equation:

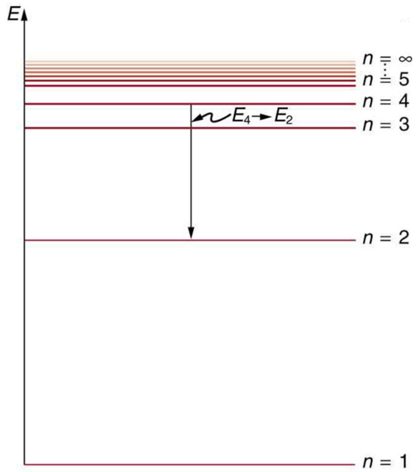
$$\Delta E = hf = E_{\text{i}} - E_{\text{f}}.$$

Here, ΔE is the change in energy between the initial and final orbits, and hf is the energy of the absorbed or emitted photon. It is quite logical (that is, expected from our everyday experience) that energy is involved in changing orbits. A blast of energy is required for the space shuttle, for example, to climb to a higher orbit. What is not expected is that atomic orbits should be quantized. This is not observed for satellites or planets, which can have any orbit given the proper energy. (See [link].)



The planetary model of the atom, as modified by Bohr, has the orbits of the electrons quantized. Only certain orbits are allowed, explaining why atomic spectra are discrete (quantized). The energy carried away from an atom by a photon comes from the electron dropping from one allowed orbit to another and is thus quantized. This is likewise true for atomic absorption of photons.

[link] shows an **energy-level diagram**, a convenient way to display energy states. In the present discussion, we take these to be the allowed energy levels of the electron. Energy is plotted vertically with the lowest or ground state at the bottom and with excited states above. Given the energies of the lines in an atomic spectrum, it is possible (although sometimes very difficult) to determine the energy levels of an atom. Energy-level diagrams are used for many systems, including molecules and nuclei. A theory of the atom or any other system must predict its energies based on the physics of the system.



An energy-level diagram plots energy vertically and is useful in visualizing the energy states of a system and the transitions between them. This diagram is for the hydrogen-atom electrons, showing a transition between two orbits having energies E_4 and E_2.

Bohr was clever enough to find a way to calculate the electron orbital energies in hydrogen. This was an important first step that has been improved upon, but it is well worth repeating here, because it does correctly describe many characteristics of hydrogen. Assuming circular orbits, Bohr proposed that the **angular momentum L of an electron in its orbit is quantized**, that is, it has only specific, discrete values. The value for L is given by the formula

Equation:

$$L = m_e \mathrm{vr}_n = n\frac{h}{2\pi}(n = 1, 2, 3, \ldots),$$

where L is the angular momentum, m_e is the electron's mass, r_n is the radius of the n th orbit, and h is Planck's constant. Note that angular momentum is $L = I\omega$. For a small object at a radius r, $I = mr^2$ and $\omega = v/r$, so that $L = \left(mr^2\right)(v/r) = mvr$. Quantization says that this value of mvr can only be equal to $h/2$, $2h/2$, $3h/2$, etc. At the time, Bohr himself did not know why angular momentum should be quantized, but using this assumption he was

able to calculate the energies in the hydrogen spectrum, something no one else had done at the time.

From Bohr's assumptions, we will now derive a number of important properties of the hydrogen atom from the classical physics we have covered in the text. We start by noting the centripetal force causing the electron to follow a circular path is supplied by the Coulomb force. To be more general, we note that this analysis is valid for any single-electron atom. So, if a nucleus has Z protons ($Z=1$ for hydrogen, 2 for helium, etc.) and only one electron, that atom is called a **hydrogen-like atom**. The spectra of hydrogen-like ions are similar to hydrogen, but shifted to higher energy by the greater attractive force between the electron and nucleus. The magnitude of the centripetal force is $m_e v^2/r_n$, while the Coulomb force is $k(\mathrm{Zq}_e)(q_e)/r_n^2$. The tacit assumption here is that the nucleus is more massive than the stationary electron, and the electron orbits about it. This is consistent with the planetary model of the atom. Equating these,

Equation:

$$k\frac{Zq_e^2}{r_n^2}=\frac{m_e v^2}{r_n}\ (\text{Coulomb} = \text{centripetal}).$$

Angular momentum quantization is stated in an earlier equation. We solve that equation for v, substitute it into the above, and rearrange the expression to obtain the radius of the orbit. This yields:

Equation:

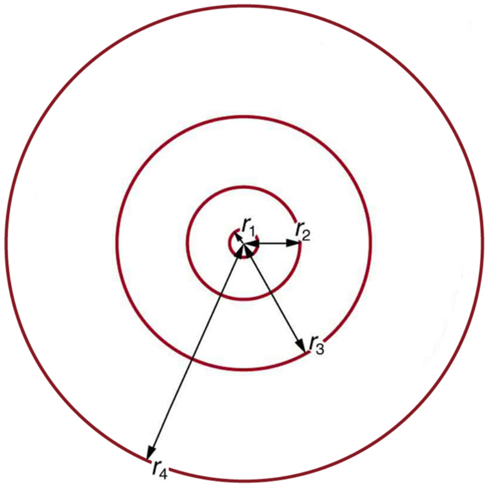
$$r_n=\frac{n^2}{Z}a_{\mathrm{B}}, \text{ for allowed orbits}(n=1{,}2{,}3,\ldots),$$

where a_{B} is defined to be the **Bohr radius**, since for the lowest orbit $(n=1)$ and for hydrogen $(Z=1)$, $r_1=a_{\mathrm{B}}$. It is left for this chapter's Problems and Exercises to show that the Bohr radius is

Equation:

$$a_{\mathrm{B}}=\frac{h^2}{4\pi^2 m_e k q_e^2}=0.529\times 10^{-10}\text{ m}.$$

These last two equations can be used to calculate the **radii of the allowed (quantized) electron orbits in any hydrogen-like atom**. It is impressive that the formula gives the correct size of hydrogen, which is measured experimentally to be very close to the Bohr radius. The earlier equation also tells us that the orbital radius is proportional to n^2, as illustrated in [link].



The allowed electron orbits in hydrogen have the radii shown. These radii were first calculated by Bohr and are given by the equation $r_n = \frac{n^2}{Z} a_{\mathrm{B}}$. The lowest orbit has the experimentally verified diameter of a hydrogen atom.

To get the electron orbital energies, we start by noting that the electron energy is the sum of its kinetic and potential energy:

Equation:

$$E_n = \mathrm{KE} + \mathrm{PE}.$$

Kinetic energy is the familiar $\mathrm{KE} = (1/2)m_e v^2$, assuming the electron is not moving at relativistic speeds. Potential energy for the electron is electrical, or $\mathrm{PE} = q_e V$, where V is the potential due to the nucleus, which looks like a point charge. The nucleus has a positive charge Zq_e ; thus, $V = \mathrm{kZq}_e / r_n$, recalling an earlier equation for the potential due to a point charge. Since the electron's charge is negative, we see that $\mathrm{PE} = -\mathrm{kZq}_e / r_n$. Entering the expressions for KE and PE, we find

Equation:

$$E_n = \frac{1}{2} m_e v^2 - k \frac{Z q_e^2}{r_n}.$$

Now we substitute r_n and v from earlier equations into the above expression for energy. Algebraic manipulation yields

Equation:

$$E_n = -\frac{Z^2}{n^2} E_0 (n = 1, 2, 3, ...)$$

for the orbital **energies of hydrogen-like atoms**. Here, E_0 is the **ground-state energy** $(n = 1)$ for hydrogen $(Z = 1)$ and is given by

Equation:

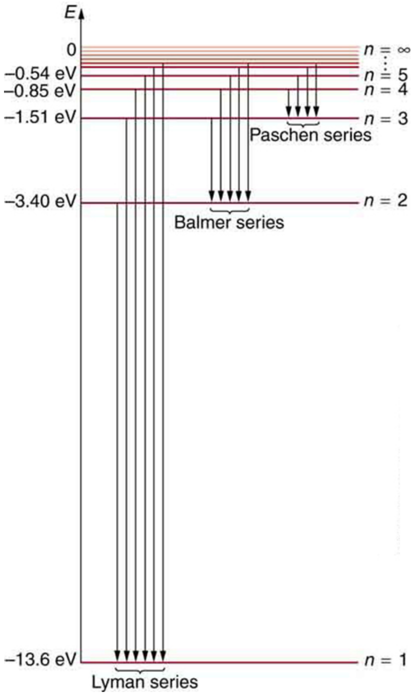
$$E_0 = \frac{2\pi^2 q_e^4 m_e k^2}{h^2} = 13.6 \text{ eV}.$$

Thus, for hydrogen,

Equation:

$$E_n = -\frac{13.6 \text{ eV}}{n^2} (n = 1, 2, 3, ...).$$

[link] shows an energy-level diagram for hydrogen that also illustrates how the various spectral series for hydrogen are related to transitions between energy levels.



Energy-level diagram for

hydrogen showing the Lyman, Balmer, and Paschen series of transitions. The orbital energies are calculated using the above equation, first derived by Bohr.

Electron total energies are negative, since the electron is bound to the nucleus, analogous to being in a hole without enough kinetic energy to escape. As n approaches infinity, the total energy becomes zero. This corresponds to a free electron with no kinetic energy, since r_n gets very large for large n, and the electric potential energy thus becomes zero. Thus, 13.6 eV is needed to ionize hydrogen (to go from –13.6 eV to 0, or unbound), an experimentally verified number. Given more energy, the electron becomes unbound with some kinetic energy. For example, giving 15.0 eV to an electron in the ground state of hydrogen strips it from the atom and leaves it with 1.4 eV of kinetic energy.

Finally, let us consider the energy of a photon emitted in a downward transition, given by the equation to be
Equation:

$$\Delta E = \text{hf} = E_{\text{i}} - E_{\text{f}}.$$

Substituting $E_n = (-13.6\text{ eV}/n^2)$, we see that
Equation:

$$\text{hf} = (13.6\text{ eV})\left(\frac{1}{n_{\text{f}}^2} - \frac{1}{n_{\text{i}}^2}\right).$$

Dividing both sides of this equation by hc gives an expression for $1/\lambda$:
Equation:

$$\frac{\text{hf}}{\text{hc}} = \frac{f}{c} = \frac{1}{\lambda} = \frac{(13.6\text{ eV})}{hc}\left(\frac{1}{n_{\text{f}}^2} - \frac{1}{n_{\text{i}}^2}\right).$$

It can be shown that
Equation:

$$\left(\frac{13.6\text{ eV}}{hc}\right) = \frac{(13.6\text{ eV})\left(1.602 \times 10^{-19}\text{ J/eV}\right)}{\left(6.626 \times 10^{-34}\text{ J}\cdot\text{s}\right)\left(2.998 \times 10^{8}\text{ m/s}\right)} = 1.097 \times 10^{7}\text{ m}^{-1} = R$$

is the **Rydberg constant**. Thus, we have used Bohr’s assumptions to derive the formula first proposed by Balmer years earlier as a recipe to fit experimental data.
Equation:

$$\frac{1}{\lambda} = R\left(\frac{1}{n_{\text{f}}^2} - \frac{1}{n_{\text{i}}^2}\right)$$

We see that Bohr’s theory of the hydrogen atom answers the question as to why this previously known formula describes the hydrogen spectrum. It is because the energy levels are proportional to $1/n^2$, where n is a non-negative integer. A downward transition releases energy, and so n_{i} must be greater than n_{f}. The various series are those where the transitions end on a certain level. For the Lyman series, $n_{\text{f}} = 1$ — that is, all the transitions end in the ground state (see also [link]). For the Balmer series, $n_{\text{f}} = 2$, or all the transitions end in the first excited state; and so on. What was once a recipe is now based in physics, and something new is emerging—angular momentum is quantized.

Triumphs and Limits of the Bohr Theory

Bohr did what no one had been able to do before. Not only did he explain the spectrum of hydrogen, he correctly calculated the size of the atom from basic physics. Some of his ideas are broadly applicable. Electron orbital energies are quantized in all atoms and molecules. Angular momentum is quantized. The electrons do not spiral into the nucleus, as expected classically (accelerated charges radiate, so that the electron orbits classically would decay quickly, and the electrons would sit on the nucleus—matter would collapse). These are major triumphs.

But there are limits to Bohr’s theory. It cannot be applied to multielectron atoms, even one as simple as a two-electron helium atom. Bohr’s model is what we call *semiclassical*. The orbits are quantized (nonclassical) but are assumed to be simple circular paths (classical). As quantum mechanics was developed, it became clear that there are no well-defined orbits; rather, there are clouds of probability. Bohr’s theory also did not explain that some spectral lines are doublets (split into two) when examined closely. We shall examine many of these aspects of quantum mechanics in more detail, but it should be kept in mind that Bohr did not fail. Rather, he made very important steps along the path to greater knowledge and laid the foundation for all of atomic physics that has since evolved.

Note:
PhET Explorations: Models of the Hydrogen Atom
How did scientists figure out the structure of atoms without looking at them? Try out different models by shooting light at the atom. Check how the prediction of the model

matches the experimental results.

https://archive.cnx.org/specials/d77cc1d0-33e4-11e6-b016-6726afecd2be/hydrogen-atom/#sim-hydrogen-atom

Section Summary

- The planetary model of the atom pictures electrons orbiting the nucleus in the way that planets orbit the sun. Bohr used the planetary model to develop the first reasonable theory of hydrogen, the simplest atom. Atomic and molecular spectra are quantized, with hydrogen spectrum wavelengths given by the formula
  Equation:
  $$\frac{1}{\lambda}=R\left(\frac{1}{n_{\mathrm{f}}^{2}}-\frac{1}{n_{\mathrm{i}}^{2}}\right),$$
  where λ is the wavelength of the emitted EM radiation and R is the Rydberg constant, which has the value
  Equation:
  $$R=1.097\times10^{7}\ \mathrm{m}^{-1}.$$
- The constants n_{i} and n_{f} are positive integers, and n_{i} must be greater than n_{f}.
- Bohr correctly proposed that the energy and radii of the orbits of electrons in atoms are quantized, with energy for transitions between orbits given by
  Equation:
  $$\Delta E=hf=E_{\mathrm{i}}-E_{\mathrm{f}},$$
  where ΔE is the change in energy between the initial and final orbits and hf is the energy of an absorbed or emitted photon. It is useful to plot orbital energies on a vertical graph called an energy-level diagram.
- Bohr proposed that the allowed orbits are circular and must have quantized orbital angular momentum given by
  Equation:
  $$L=m_{e}vr_{n}=n\frac{h}{2\pi}(n=1,2,3\ldots),$$
  where L is the angular momentum, r_n is the radius of the $n\mathrm{th}$ orbit, and h is Planck's constant. For all one-electron (hydrogen-like) atoms, the radius of an orbit is given by
  Equation:
  $$r_{n}=\frac{n^{2}}{Z}a_{\mathrm{B}}(\text{allowed orbits } n=1,2,3,\ldots),$$

Z is the atomic number of an element (the number of electrons is has when neutral) and a_{B} is defined to be the Bohr radius, which is
Equation:

$$a_{\mathrm{B}} = \frac{h^2}{4\pi^2 m_e k q_e^2} = 0.529 \times 10^{-10}\ \mathrm{m}.$$

- Furthermore, the energies of hydrogen-like atoms are given by
  Equation:

  $$E_n = -\frac{Z^2}{n^2} E_0 (n = 1, 2, 3 \ldots),$$

  where E_0 is the ground-state energy and is given by
  Equation:

  $$E_0 = \frac{2\pi^2 q_e^4 m_e k^2}{h^2} = 13.6\ \mathrm{eV}.$$

  Thus, for hydrogen,
  Equation:

  $$E_n = -\frac{13.6\ \mathrm{eV}}{n^2} (n, =, 1, 2, 3 \ldots).$$

- The Bohr Theory gives accurate values for the energy levels in hydrogen-like atoms, but it has been improved upon in several respects.

Conceptual Questions

Exercise:

Problem:

How do the allowed orbits for electrons in atoms differ from the allowed orbits for planets around the sun? Explain how the correspondence principle applies here.

Exercise:

Problem:

Explain how Bohr’s rule for the quantization of electron orbital angular momentum differs from the actual rule.

Exercise:

Problem:

What is a hydrogen-like atom, and how are the energies and radii of its electron orbits related to those in hydrogen?

Problems & Exercises

Exercise:

Problem:

By calculating its wavelength, show that the first line in the Lyman series is UV radiation.

Solution:

$\frac{1}{\lambda} = R\left(\frac{1}{n_\mathrm{f}^2} - \frac{1}{n_\mathrm{i}^2}\right) \Rightarrow \lambda = \frac{1}{R}\left[\frac{(n_\mathrm{i} \cdot n_\mathrm{f})^2}{n_\mathrm{i}^2 - n_\mathrm{f}^2}\right]; n_\mathrm{i} = 2, n_\mathrm{f} = 1,$ so that

$\lambda = \left(\frac{\mathrm{m}}{1.097 \times 10^7}\right)\left[\frac{(2 \times 1)^2}{2^2 - 1^2}\right] = 1.22 \times 10^{-7}\ \mathrm{m} = 122\ \mathrm{nm}$, which is UV radiation.

Exercise:

Problem:

Find the wavelength of the third line in the Lyman series, and identify the type of EM radiation.

Exercise:

Problem:

Look up the values of the quantities in $a_\mathrm{B} = \frac{h^2}{4\pi^2 m_e k q_e^2}$, and verify that the Bohr radius a_B is 0.529×10^{-10} m.

Solution:

$$a_\mathrm{B} = \frac{h^2}{4\pi^2 m_e k Z q_e^2} = \frac{(6.626 \times 10^{-34}\ \mathrm{J \cdot s})^2}{4\pi^2 (9.109 \times 10^{-31}\ \mathrm{kg})(8.988 \times 10^9\ \mathrm{N \cdot m^2/C^2})(1)(1.602 \times 10^{-19}\ \mathrm{C})^2} = 0.529 \times 10^{-10}\ \mathrm{m}$$

Exercise:

Problem: Verify that the ground state energy E_0 is 13.6 eV by using $E_0 = \frac{2\pi^2 q_e^4 m_e k^2}{h^2}$.

Exercise:

Problem:

If a hydrogen atom has its electron in the $n = 4$ state, how much energy in eV is needed to ionize it?

Solution:

0.850 eV

Exercise:

Problem:

A hydrogen atom in an excited state can be ionized with less energy than when it is in its ground state. What is n for a hydrogen atom if 0.850 eV of energy can ionize it?

Exercise:

Problem:

Find the radius of a hydrogen atom in the $n = 2$ state according to Bohr's theory.

Solution:

$2.12 \times 10^{-10}\ \text{m}$

Exercise:

Problem:

Show that $(13.6\ \text{eV})/\text{hc} = 1.097 \times 10^{7}\ \text{m} = R$ (Rydberg's constant), as discussed in the text.

Exercise:

Problem:

What is the smallest-wavelength line in the Balmer series? Is it in the visible part of the spectrum?

Solution:

365 nm

It is in the ultraviolet.

Exercise:

Problem:

Show that the entire Paschen series is in the infrared part of the spectrum. To do this, you only need to calculate the shortest wavelength in the series.

Exercise:

Problem:

Do the Balmer and Lyman series overlap? To answer this, calculate the shortest-wavelength Balmer line and the longest-wavelength Lyman line.

Solution:

No overlap

365 nm

122 nm

Exercise:

Problem:

(a) Which line in the Balmer series is the first one in the UV part of the spectrum?

(b) How many Balmer series lines are in the visible part of the spectrum?

(c) How many are in the UV?

Exercise:

Problem:

A wavelength of $4.653\ \mu\text{m}$ is observed in a hydrogen spectrum for a transition that ends in the $n_{\text{f}} = 5$ level. What was n_{i} for the initial level of the electron?

Solution:

7

Exercise:

Problem:

A singly ionized helium ion has only one electron and is denoted He^{+}. What is the ion's radius in the ground state compared to the Bohr radius of hydrogen atom?

Exercise:

Problem:

A beryllium ion with a single electron (denoted Be^{3+}) is in an excited state with radius the same as that of the ground state of hydrogen.

(a) What is n for the Be^{3+} ion?

(b) How much energy in eV is needed to ionize the ion from this excited state?

Solution:

(a) 2

(b) 54.4 eV

Exercise:

Problem:

Atoms can be ionized by thermal collisions, such as at the high temperatures found in the solar corona. One such ion is C^{+5}, a carbon atom with only a single electron.

(a) By what factor are the energies of its hydrogen-like levels greater than those of hydrogen?

(b) What is the wavelength of the first line in this ion's Paschen series?

(c) What type of EM radiation is this?

Exercise:

Problem:

Verify Equations $r_n = \frac{n^2}{Z} a_{\mathrm{B}}$ and $a_{\mathrm{B}} = \frac{h^2}{4\pi^2 m_e k q_e^2} = 0.529 \times 10^{-10}\ \mathrm{m}$ using the approach stated in the text. That is, equate the Coulomb and centripetal forces and then insert an expression for velocity from the condition for angular momentum quantization.

Solution:

$\frac{kZq_e^2}{r_n^2} = \frac{m_e V^2}{r_n}$, so that $r_n = \frac{kZq_e^2}{m_e V^2} = \frac{kZq_e^2}{m_e} \frac{1}{V^2}$. From the equation $m_e v r_n = n \frac{h}{2\pi}$, we can substitute for the velocity, giving: $r_n = \frac{kZq_e^2}{m_e} \cdot \frac{4\pi^2 m_e^2 r_n^2}{n^2 h^2}$ so that $r_n = \frac{n^2}{Z} \frac{h^2}{4\pi^2 m_e k q_e^2} = \frac{n^2}{Z} a_{\mathrm{B}}$, where $a_{\mathrm{B}} = \frac{h^2}{4\pi^2 m_e k q_e^2}$.

Exercise:

Problem:

The wavelength of the four Balmer series lines for hydrogen are found to be 410.3, 434.2, 486.3, and 656.5 nm. What average percentage difference is found between these wavelength numbers and those predicted by $\frac{1}{\lambda} = R\left(\frac{1}{n_{\mathrm{f}}^2} - \frac{1}{n_{\mathrm{i}}^2}\right)$? It is amazing how well a simple formula (disconnected originally from theory) could duplicate this phenomenon.

Glossary

hydrogen spectrum wavelengths
: the wavelengths of visible light from hydrogen; can be calculated by $\frac{1}{\lambda} = R\left(\frac{1}{n_{\mathrm{f}}^2} - \frac{1}{n_{\mathrm{i}}^2}\right)$

Rydberg constant
: a physical constant related to the atomic spectra with an established value of $1.097 \times 10^7\ \mathrm{m}^{-1}$

double-slit interference
: an experiment in which waves or particles from a single source impinge upon two slits so that the resulting interference pattern may be observed

energy-level diagram
: a diagram used to analyze the energy level of electrons in the orbits of an atom

Bohr radius
: the mean radius of the orbit of an electron around the nucleus of a hydrogen atom in its ground state

hydrogen-like atom
: any atom with only a single electron

energies of hydrogen-like atoms
: Bohr formula for energies of electron states in hydrogen-like atoms: $E_n = -\frac{Z^2}{n^2}E_0(n = 1, 2, 3,\ \ldots)$

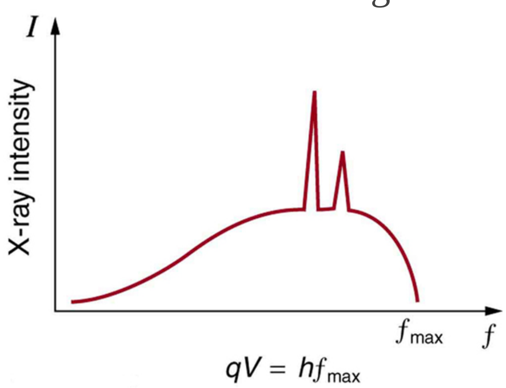
X Rays: Atomic Origins and Applications

- Define x-ray tube and its spectrum.
- Show the x-ray characteristic energy.
- Specify the use of x rays in medical observations.
- Explain the use of x rays in CT scanners in diagnostics.

Each type of atom (or element) has its own characteristic electromagnetic spectrum. **X rays** lie at the high-frequency end of an atom's spectrum and are characteristic of the atom as well. In this section, we explore characteristic x rays and some of their important applications.

We have previously discussed x rays as a part of the electromagnetic spectrum in Photon Energies and the Electromagnetic Spectrum. That module illustrated how an x-ray tube (a specialized CRT) produces x rays. Electrons emitted from a hot filament are accelerated with a high voltage, gaining significant kinetic energy and striking the anode.

There are two processes by which x rays are produced in the anode of an x-ray tube. In one process, the deceleration of electrons produces x rays, and these x rays are called *bremsstrahlung*, or braking radiation. The second process is atomic in nature and produces *characteristic x rays*, so called because they are characteristic of the anode material. The x-ray spectrum in [link] is typical of what is produced by an x-ray tube, showing a broad curve of bremsstrahlung radiation with characteristic x-ray peaks on it.



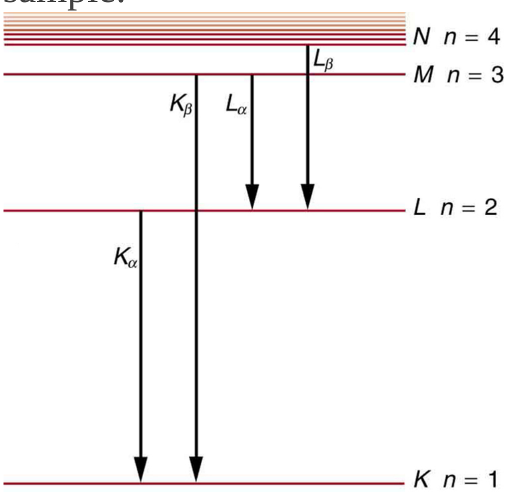
X-ray spectrum obtained when energetic electrons strike a material, such as

in the anode of a CRT. The smooth part of the spectrum is bremsstrahlung radiation, while the peaks are characteristic of the anode material. A different anode material would have characteristic x-ray peaks at different frequencies.

The spectrum in [link] is collected over a period of time in which many electrons strike the anode, with a variety of possible outcomes for each hit. The broad range of x-ray energies in the bremsstrahlung radiation indicates that an incident electron's energy is not usually converted entirely into photon energy. The highest-energy x ray produced is one for which all of the electron's energy was converted to photon energy. Thus the accelerating voltage and the maximum x-ray energy are related by conservation of energy. Electric potential energy is converted to kinetic energy and then to photon energy, so that $E_{\text{max}} = hf_{\text{max}} = q_e V$. Units of electron volts are convenient. For example, a 100-kV accelerating voltage produces x-ray photons with a maximum energy of 100 keV.

Some electrons excite atoms in the anode. Part of the energy that they deposit by collision with an atom results in one or more of the atom's inner electrons being knocked into a higher orbit or the atom being ionized. When the anode's atoms de-excite, they emit characteristic electromagnetic radiation. The most energetic of these are produced when an inner-shell vacancy is filled—that is, when an $n = 1$ or $n = 2$ shell electron has been excited to a higher level, and another electron falls into the vacant spot. A *characteristic x ray* (see Photon Energies and the Electromagnetic Spectrum) is electromagnetic (EM) radiation emitted by an atom when an inner-shell vacancy is filled. [link] shows a representative energy-level diagram that illustrates the labeling of characteristic x rays. X rays created

when an electron falls into an $n = 1$ shell vacancy are called K_α when they come from the next higher level; that is, an $n = 2$ to $n = 1$ transition. The labels $K, L, M, \ldots$ come from the older alphabetical labeling of shells starting with K rather than using the principal quantum numbers 1, 2, 3, $\ldots$. A more energetic K_β x ray is produced when an electron falls into an $n = 1$ shell vacancy from the $n = 3$ shell; that is, an $n = 3$ to $n = 1$ transition. Similarly, when an electron falls into the $n = 2$ shell from the $n = 3$ shell, an L_α x ray is created. The energies of these x rays depend on the energies of electron states in the particular atom and, thus, are characteristic of that element: every element has it own set of x-ray energies. This property can be used to identify elements, for example, to find trace (small) amounts of an element in an environmental or biological sample.



A characteristic x ray is emitted when an electron fills an inner-shell vacancy, as shown for several transitions in this approximate energy level diagram for a multiple-electron atom. Characteristic x rays are labeled according to the shell that had the vacancy and the shell from which

the electron came. A K_α x ray, for example, is produced when an electron coming from the $n = 2$ shell fills the $n = 1$ shell vacancy.

Example:
Characteristic X-Ray Energy

Calculate the approximate energy of a K_α x ray from a tungsten anode in an x-ray tube.

Strategy

How do we calculate energies in a multiple-electron atom? In the case of characteristic x rays, the following approximate calculation is reasonable. Characteristic x rays are produced when an inner-shell vacancy is filled. Inner-shell electrons are nearer the nucleus than others in an atom and thus feel little net effect from the others. This is similar to what happens inside a charged conductor, where its excess charge is distributed over the surface so that it produces no electric field inside. It is reasonable to assume the inner-shell electrons have hydrogen-like energies, as given by $E_n = -\frac{Z^2}{n^2} E_0 (n = 1, 2, 3, ...)$. As noted, a K_α x ray is produced by an $n = 2$ to $n = 1$ transition. Since there are two electrons in a filled K shell, a vacancy would leave one electron, so that the effective charge would be $Z - 1$ rather than Z. For tungsten, $Z = 74$, so that the effective charge is 73.

Solution

$E_n = -\frac{Z^2}{n^2} E_0 (n = 1, 2, 3, ...)$ gives the orbital energies for hydrogen-like atoms to be $E_n = -(Z^2/n^2)E_0$, where $E_0 = 13.6$ eV. As noted, the effective Z is 73. Now the K_α x-ray energy is given by

Equation:

$$E_{K_\alpha} = \Delta E = E_\mathrm{i} - E_\mathrm{f} = E_2 - E_1,$$

where

Equation:

$$E_1 = -\frac{Z^2}{1^2}E_0 = -\frac{73^2}{1}\left(13.6\text{ eV}\right) = -72.5\text{ keV}$$

and

Equation:

$$E_2 = -\frac{Z^2}{2^2}E_0 = -\frac{73^2}{4}\left(13.6\text{ eV}\right) = -18.1\text{ keV}.$$

Thus,

Equation:

$$E_{K_\alpha} = -18.1\text{ keV} - (-72.5\text{ keV}) = 54.4\text{ keV}.$$

Discussion

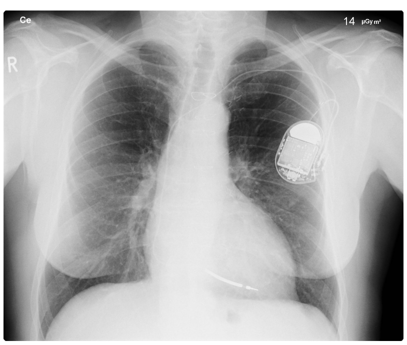
This large photon energy is typical of characteristic x rays from heavy elements. It is large compared with other atomic emissions because it is produced when an inner-shell vacancy is filled, and inner-shell electrons are tightly bound. Characteristic x ray energies become progressively larger for heavier elements because their energy increases approximately as Z^2. Significant accelerating voltage is needed to create these inner-shell vacancies. In the case of tungsten, at least 72.5 kV is needed, because other shells are filled and you cannot simply bump one electron to a higher filled shell. Tungsten is a common anode material in x-ray tubes; so much of the energy of the impinging electrons is absorbed, raising its temperature, that a high-melting-point material like tungsten is required.

Medical and Other Diagnostic Uses of X-rays

All of us can identify diagnostic uses of x-ray photons. Among these are the universal dental and medical x rays that have become an essential part of medical diagnostics. (See [link] and [link].) X rays are also used to inspect

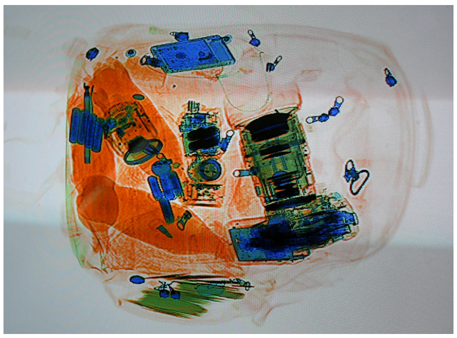
our luggage at airports, as shown in [link], and for early detection of cracks in crucial aircraft components. An x ray is not only a noun meaning high-energy photon, it also is an image produced by x rays, and it has been made into a familiar verb—to be x-rayed.

An x-ray image reveals fillings in a person's teeth. (credit: Dmitry G, Wikimedia Commons)



This x-ray image of a person's chest shows many details, including an artificial pacemaker. (credit: Sunzi99,

Wikimedia
Commons)



This x-ray image shows the contents of a piece of luggage. The denser the material, the darker the shadow. (credit: IDuke, Wikimedia Commons)

The most common x-ray images are simple shadows. Since x-ray photons have high energies, they penetrate materials that are opaque to visible light. The more energy an x-ray photon has, the more material it will penetrate. So an x-ray tube may be operated at 50.0 kV for a chest x ray, whereas it may need to be operated at 100 kV to examine a broken leg in a cast. The depth of penetration is related to the density of the material as well as to the energy of the photon. The denser the material, the fewer x-ray photons get through and the darker the shadow. Thus x rays excel at detecting breaks in bones and in imaging other physiological structures, such as some tumors, that differ in density from surrounding material. Because of their high photon energy, x rays produce significant ionization in materials and

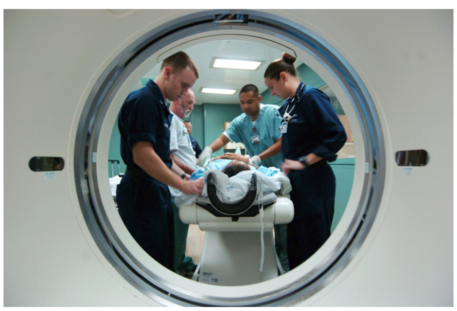
damage cells in biological organisms. Modern uses minimize exposure to the patient and eliminate exposure to others. Biological effects of x rays will be explored in the next chapter along with other types of ionizing radiation such as those produced by nuclei.

As the x-ray energy increases, the Compton effect (see Photon Momentum) becomes more important in the attenuation of the x rays. Here, the x ray scatters from an outer electron shell of the atom, giving the ejected electron some kinetic energy while losing energy itself. The probability for attenuation of the x rays depends upon the number of electrons present (the material's density) as well as the thickness of the material. Chemical composition of the medium, as characterized by its atomic number Z, is not important here. Low-energy x rays provide better contrast (sharper images). However, due to greater attenuation and less scattering, they are more absorbed by thicker materials. Greater contrast can be achieved by injecting a substance with a large atomic number, such as barium or iodine. The structure of the part of the body that contains the substance (e.g., the gastro-intestinal tract or the abdomen) can easily be seen this way.

Breast cancer is the second-leading cause of death among women worldwide. Early detection can be very effective, hence the importance of x-ray diagnostics. A mammogram cannot diagnose a malignant tumor, only give evidence of a lump or region of increased density within the breast. X-ray absorption by different types of soft tissue is very similar, so contrast is difficult; this is especially true for younger women, who typically have denser breasts. For older women who are at greater risk of developing breast cancer, the presence of more fat in the breast gives the lump or tumor more contrast. MRI (Magnetic resonance imaging) has recently been used as a supplement to conventional x rays to improve detection and eliminate false positives. The subject's radiation dose from x rays will be treated in a later chapter.

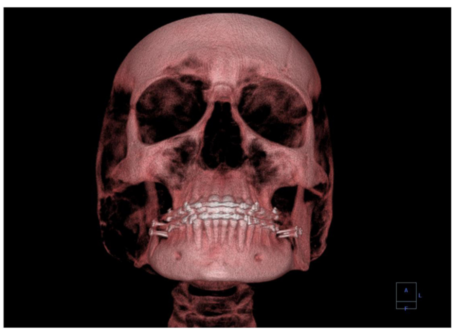
A standard x ray gives only a two-dimensional view of the object. Dense bones might hide images of soft tissue or organs. If you took another x ray from the side of the person (the first one being from the front), you would gain additional information. While shadow images are sufficient in many applications, far more sophisticated images can be produced with modern

technology. [link] shows the use of a computed tomography (CT) scanner, also called computed axial tomography (CAT) scanner. X rays are passed through a narrow section (called a slice) of the patient's body (or body part) over a range of directions. An array of many detectors on the other side of the patient registers the x rays. The system is then rotated around the patient and another image is taken, and so on. The x-ray tube and detector array are mechanically attached and so rotate together. Complex computer image processing of the relative absorption of the x rays along different directions produces a highly-detailed image. Different slices are taken as the patient moves through the scanner on a table. Multiple images of different slices can also be computer analyzed to produce three-dimensional information, sometimes enhancing specific types of tissue, as shown in [link]. G. Hounsfield (UK) and A. Cormack (US) won the Nobel Prize in Medicine in 1979 for their development of computed tomography.



A patient being positioned in a CT scanner aboard the hospital ship USNS Mercy. The CT scanner passes x rays through slices of the patient's body (or body part) over a range of directions. The relative absorption of the x rays along different

directions is computer analyzed to produce highly detailed images. Three-dimensional information can be obtained from multiple slices. (credit: Rebecca Moat, U.S. Navy)



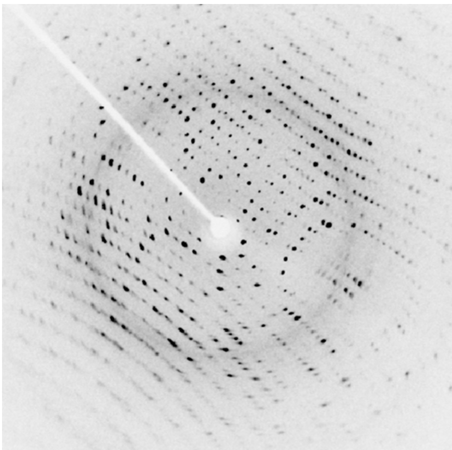
This three-dimensional image of a skull was produced by computed tomography, involving analysis of several x-ray slices of the head. (credit: Emailshankar, Wikimedia Commons)

X-Ray Diffraction and Crystallography

Since x-ray photons are very energetic, they have relatively short wavelengths. For example, the 54.4-keV K_α x ray of [link] has a

wavelength $\lambda = \mathrm{hc}/E = 0.0228\ \mathrm{nm}$. Thus, typical x-ray photons act like rays when they encounter macroscopic objects, like teeth, and produce sharp shadows; however, since atoms are on the order of 0.1 nm in size, x rays can be used to detect the location, shape, and size of atoms and molecules. The process is called **x-ray diffraction**, because it involves the diffraction and interference of x rays to produce patterns that can be analyzed for information about the structures that scattered the x rays. Perhaps the most famous example of x-ray diffraction is the discovery of the double-helix structure of DNA in 1953 by an international team of scientists working at the Cavendish Laboratory—American James Watson, Englishman Francis Crick, and New Zealand–born Maurice Wilkins. Using x-ray diffraction data produced by Rosalind Franklin, they were the first to discern the structure of DNA that is so crucial to life. For this, Watson, Crick, and Wilkins were awarded the 1962 Nobel Prize in Physiology or Medicine. There is much debate and controversy over the issue that Rosalind Franklin was not included in the prize.

[link] shows a diffraction pattern produced by the scattering of x rays from a crystal. This process is known as x-ray crystallography because of the information it can yield about crystal structure, and it was the type of data Rosalind Franklin supplied to Watson and Crick for DNA. Not only do x rays confirm the size and shape of atoms, they give information on the atomic arrangements in materials. For example, current research in high-temperature superconductors involves complex materials whose lattice arrangements are crucial to obtaining a superconducting material. These can be studied using x-ray crystallography.



X-ray diffraction from the crystal of a protein, hen egg lysozyme, produced this interference pattern. Analysis of the pattern yields information about the structure of the protein. (credit: Del45, Wikimedia Commons)

Historically, the scattering of x rays from crystals was used to prove that x rays are energetic EM waves. This was suspected from the time of the discovery of x rays in 1895, but it was not until 1912 that the German Max von Laue (1879–1960) convinced two of his colleagues to scatter x rays from crystals. If a diffraction pattern is obtained, he reasoned, then the x rays must be waves, and their wavelength could be determined. (The spacing of atoms in various crystals was reasonably well known at the time, based on good values for Avogadro’s number.) The experiments were convincing, and the 1914 Nobel Prize in Physics was given to von Laue for his suggestion leading to the proof that x rays are EM waves. In 1915, the unique father-and-son team of Sir William Henry Bragg and his son Sir William Lawrence Bragg were awarded a joint Nobel Prize for inventing

the x-ray spectrometer and the then-new science of x-ray analysis. The elder Bragg had migrated to Australia from England just after graduating in mathematics. He learned physics and chemistry during his career at the University of Adelaide. The younger Bragg was born in Adelaide but went back to the Cavendish Laboratories in England to a career in x-ray and neutron crystallography; he provided support for Watson, Crick, and Wilkins for their work on unraveling the mysteries of DNA and to Max Perutz for his 1962 Nobel Prize-winning work on the structure of hemoglobin. Here again, we witness the enabling nature of physics—establishing instruments and designing experiments as well as solving mysteries in the biomedical sciences.

Certain other uses for x rays will be studied in later chapters. X rays are useful in the treatment of cancer because of the inhibiting effect they have on cell reproduction. X rays observed coming from outer space are useful in determining the nature of their sources, such as neutron stars and possibly black holes. Created in nuclear bomb explosions, x rays can also be used to detect clandestine atmospheric tests of these weapons. X rays can cause excitations of atoms, which then fluoresce (emitting characteristic EM radiation), making x-ray-induced fluorescence a valuable analytical tool in a range of fields from art to archaeology.

Section Summary

- X rays are relatively high-frequency EM radiation. They are produced by transitions between inner-shell electron levels, which produce x rays characteristic of the atomic element, or by decelerating electrons.
- X rays have many uses, including medical diagnostics and x-ray diffraction.

Conceptual Questions

Exercise:

Problem:

Explain why characteristic x rays are the most energetic in the EM emission spectrum of a given element.

Exercise:

Problem:

Why does the energy of characteristic x rays become increasingly greater for heavier atoms?

Exercise:

Problem:

Observers at a safe distance from an atmospheric test of a nuclear bomb feel its heat but receive none of its copious x rays. Why is air opaque to x rays but transparent to infrared?

Exercise:

Problem:

Lasers are used to burn and read CDs. Explain why a laser that emits blue light would be capable of burning and reading more information than one that emits infrared.

Exercise:

Problem:

Crystal lattices can be examined with x rays but not UV. Why?

Exercise:

Problem:

CT scanners do not detect details smaller than about 0.5 mm. Is this limitation due to the wavelength of x rays? Explain.

Problem Exercises

Exercise:

Problem:

(a) What is the shortest-wavelength x-ray radiation that can be generated in an x-ray tube with an applied voltage of 50.0 kV? (b) Calculate the photon energy in eV. (c) Explain the relationship of the photon energy to the applied voltage.

Solution:

(a) 0.248×10^{-10} m

(b) 50.0 keV

(c) The photon energy is simply the applied voltage times the electron charge, so the value of the voltage in volts is the same as the value of the energy in electron volts.

Exercise:

Problem:

A color television tube also generates some x rays when its electron beam strikes the screen. What is the shortest wavelength of these x rays, if a 30.0-kV potential is used to accelerate the electrons? (Note that TVs have shielding to prevent these x rays from exposing viewers.)

Exercise:

Problem:

An x ray tube has an applied voltage of 100 kV. (a) What is the most energetic x-ray photon it can produce? Express your answer in electron volts and joules. (b) Find the wavelength of such an X–ray.

Solution:

(a) 100×10^{3} eV, 1.60×10^{-14} J

(b) 0.124×10^{-10} m

Exercise:

Problem:

The maximum characteristic x-ray photon energy comes from the capture of a free electron into a K shell vacancy. What is this photon energy in keV for tungsten, assuming the free electron has no initial kinetic energy?

Exercise:

Problem:

What are the approximate energies of the K_α and K_β x rays for copper?

Solution:

(a) 8.00 keV

(b) 9.48 keV

Glossary

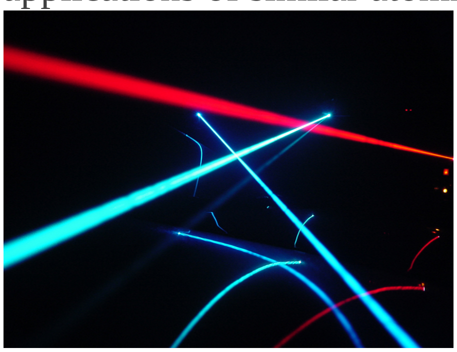
x rays
: a form of electromagnetic radiation

x-ray diffraction
: a technique that provides the detailed information about crystallographic structure of natural and manufactured materials

Applications of Atomic Excitations and De-Excitations

- Define and discuss fluorescence.
- Define metastable.
- Describe how laser emission is produced.
- Explain population inversion.
- Define and discuss holography.

Many properties of matter and phenomena in nature are directly related to atomic energy levels and their associated excitations and de-excitations. The color of a rose, the output of a laser, and the transparency of air are but a few examples. (See [link].) While it may not appear that glow-in-the-dark pajamas and lasers have much in common, they are in fact different applications of similar atomic de-excitations.



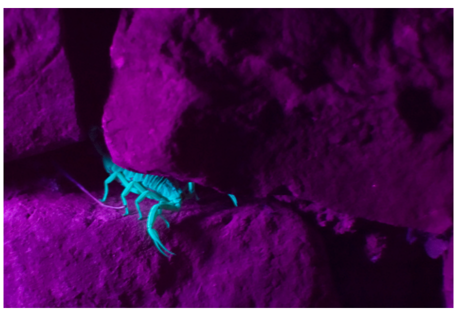
Light from a laser is based on a particular type of atomic de-excitation. (credit: Jeff Keyzer)

The color of a material is due to the ability of its atoms to absorb certain wavelengths while reflecting or reemitting others. A simple red material, for example a tomato, absorbs all visible wavelengths except red. This is because the atoms of its hydrocarbon pigment (lycopene) have levels separated by a variety of energies corresponding to all visible photon energies except red. Air is another interesting example. It is transparent to visible light, because there are few energy levels that visible photons can

excite in air molecules and atoms. Visible light, thus, cannot be absorbed. Furthermore, visible light is only weakly scattered by air, because visible wavelengths are so much greater than the sizes of the air molecules and atoms. Light must pass through kilometers of air to scatter enough to cause red sunsets and blue skies.

Fluorescence and Phosphorescence

The ability of a material to emit various wavelengths of light is similarly related to its atomic energy levels. [link] shows a scorpion illuminated by a UV lamp, sometimes called a black light. Some rocks also glow in black light, the particular colors being a function of the rock's mineral composition. Black lights are also used to make certain posters glow.



Objects glow in the visible spectrum when illuminated by an ultraviolet (black) light. Emissions are characteristic of the mineral involved, since they are related to its energy levels. In the case of scorpions, proteins near the surface of their skin give off the characteristic blue

glow. This is a colorful example of fluorescence in which excitation is induced by UV radiation while de-excitation occurs in the form of visible light. (credit: Ken Bosma, Flickr)

In the fluorescence process, an atom is excited to a level several steps above its ground state by the absorption of a relatively high-energy UV photon. This is called **atomic excitation**. Once it is excited, the atom can de-excite in several ways, one of which is to re-emit a photon of the same energy as excited it, a single step back to the ground state. This is called **atomic de-excitation**. All other paths of de-excitation involve smaller steps, in which lower-energy (longer wavelength) photons are emitted. Some of these may be in the visible range, such as for the scorpion in [link]. **Fluorescence** is defined to be any process in which an atom or molecule, excited by a photon of a given energy, and de-excites by emission of a lower-energy photon.

Fluorescence can be induced by many types of energy input. Fluorescent paint, dyes, and even soap residues in clothes make colors seem brighter in sunlight by converting some UV into visible light. X rays can induce fluorescence, as is done in x-ray fluoroscopy to make brighter visible images. Electric discharges can induce fluorescence, as in so-called neon lights and in gas-discharge tubes that produce atomic and molecular spectra. Common fluorescent lights use an electric discharge in mercury vapor to cause atomic emissions from mercury atoms. The inside of a fluorescent light is coated with a fluorescent material that emits visible light over a broad spectrum of wavelengths. By choosing an appropriate coating, fluorescent lights can be made more like sunlight or like the reddish glow of candlelight, depending on needs. Fluorescent lights are more efficient in converting electrical energy into visible light than incandescent filaments

(about four times as efficient), the blackbody radiation of which is primarily in the infrared due to temperature limitations.

This atom is excited to one of its higher levels by absorbing a UV photon. It can de-excite in a single step, re-emitting a photon of the same energy, or in several steps. The process is called fluorescence if the atom de-excites in smaller steps, emitting energy different from that which excited it. Fluorescence can be induced by a variety of energy inputs, such as UV, x-rays, and electrical discharge.

The spectacular Waitomo caves on North Island in New Zealand provide a natural habitat for glow-worms. The glow-worms hang up to 70 silk threads of about 30 or 40 cm each to trap prey that fly towards them in the dark. The fluorescence process is very efficient, with nearly 100% of the energy input turning into light. (In comparison, fluorescent lights are about 20% efficient.)

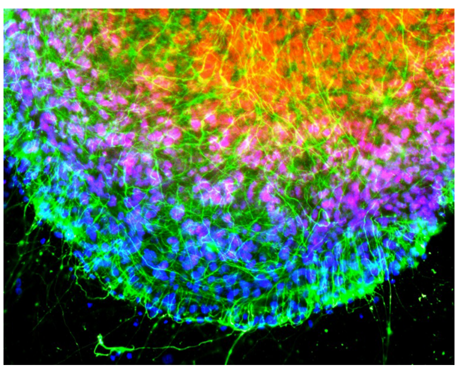
Fluorescence has many uses in biology and medicine. It is commonly used to label and follow a molecule within a cell. Such tagging allows one to study the structure of DNA and proteins. Fluorescent dyes and antibodies are usually used to tag the molecules, which are then illuminated with UV light and their emission of visible light is observed. Since the fluorescence of each element is characteristic, identification of elements within a sample can be done this way.

[link] shows a commonly used fluorescent dye called fluorescein. Below that, [link] reveals the diffusion of a fluorescent dye in water by observing it under UV light.

Fluorescein, shown here in powder form, is used to dye laboratory samples. (credit: Benjah-bmm27, Wikimedia Commons)

Here, fluorescent powder is added to a beaker of water. The mixture gives off a bright glow under ultraviolet light. (credit: Bricksnite, Wikimedia Commons)

Note:
Nano-Crystals
Recently, a new class of fluorescent materials has appeared—“nano-crystals.” These are single-crystal molecules less than 100 nm in size. The smallest of these are called “quantum dots.” These semiconductor indicators are very small (2–6 nm) and provide improved brightness. They also have the advantage that all colors can be excited with the same incident wavelength. They are brighter and more stable than organic dyes and have a longer lifetime than conventional phosphors. They have become an excellent tool for long-term studies of cells, including migration and morphology. ([link].)



Microscopic image of chicken cells using nano-crystals of a fluorescent dye. Cell nuclei exhibit blue fluorescence while neurofilaments exhibit green. (credit: Weerapong Prasongchean, Wikimedia Commons)

Once excited, an atom or molecule will usually spontaneously de-excite quickly. (The electrons raised to higher levels are attracted to lower ones by the positive charge of the nucleus.) Spontaneous de-excitation has a very short mean lifetime of typically about 10^{-8} s. However, some levels have significantly longer lifetimes, ranging up to milliseconds to minutes or even hours. These energy levels are inhibited and are slow in de-exciting because their quantum numbers differ greatly from those of available lower levels. Although these level lifetimes are short in human terms, they are many orders of magnitude longer than is typical and, thus, are said to be **metastable**, meaning relatively stable. **Phosphorescence** is the de-excitation of a metastable state. Glow-in-the-dark materials, such as luminous dials on some watches and clocks and on children’s toys and pajamas, are made of phosphorescent substances. Visible light excites the atoms or molecules to metastable states that decay slowly, releasing the stored excitation energy partially as visible light. In some ceramics, atomic excitation energy can be frozen in after the ceramic has cooled from its firing. It is very slowly released, but the ceramic can be induced to phosphoresce by heating—a process called “thermoluminescence.” Since the release is slow, thermoluminescence can be used to date antiquities. The less light emitted, the older the ceramic. (See [link].)

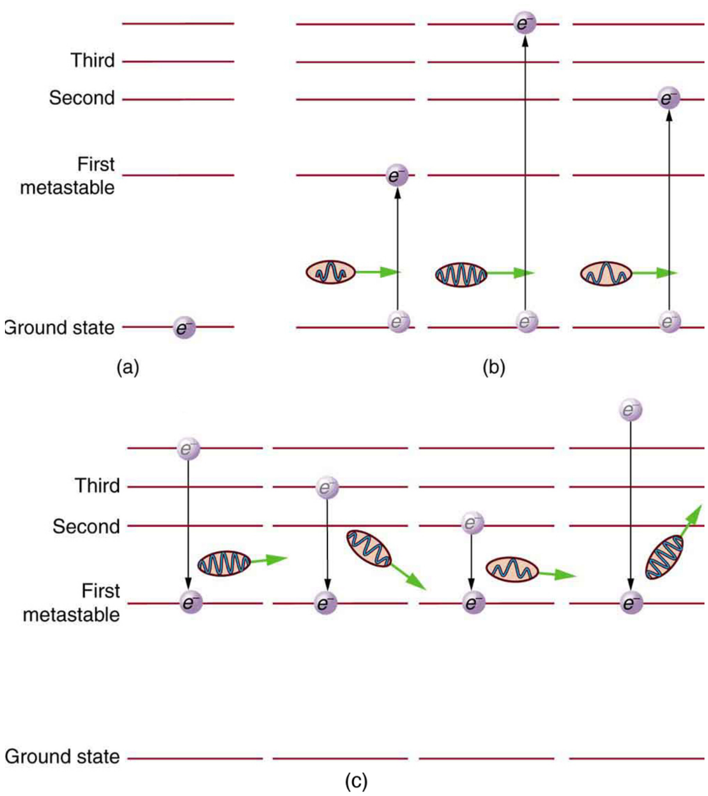
Atoms frozen in an excited state when this Chinese ceramic figure was fired can be stimulated to de-excite and emit EM radiation by heating a sample of the ceramic—a process called thermoluminescence. Since the states slowly de-excite over centuries, the amount of thermoluminescence decreases with age, making it possible to use this effect to date and authenticate antiquities. This figure dates from the 11th century. (credit: Vassil, Wikimedia Commons)

Lasers

Lasers today are commonplace. Lasers are used to read bar codes at stores and in libraries, laser shows are staged for entertainment, laser printers produce high-quality images at relatively low cost, and lasers send prodigious numbers of telephone messages through optical fibers. Among other things, lasers are also employed in surveying, weapons guidance,

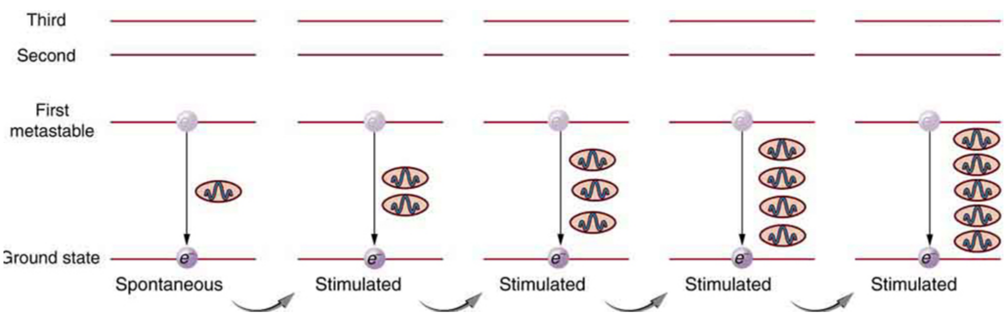
tumor eradication, retinal welding, and for reading music CDs and computer CD-ROMs.

Why do lasers have so many varied applications? The answer is that lasers produce single-wavelength EM radiation that is also very coherent—that is, the emitted photons are in phase. Laser output can, thus, be more precisely manipulated than incoherent mixed-wavelength EM radiation from other sources. The reason laser output is so pure and coherent is based on how it is produced, which in turn depends on a metastable state in the lasing material. Suppose a material had the energy levels shown in [link]. When energy is put into a large collection of these atoms, electrons are raised to all possible levels. Most return to the ground state in less than about 10^{-8} s, but those in the metastable state linger. This includes those electrons originally excited to the metastable state and those that fell into it from above. It is possible to get a majority of the atoms into the metastable state, a condition called a **population inversion**.



(a) Energy-level diagram for an atom showing the first few states, one of which is metastable. (b) Massive energy input excites atoms to a variety of states. (c) Most states decay quickly, leaving electrons only in the metastable and ground state. If a majority of electrons are in the metastable state, a population inversion has been achieved.

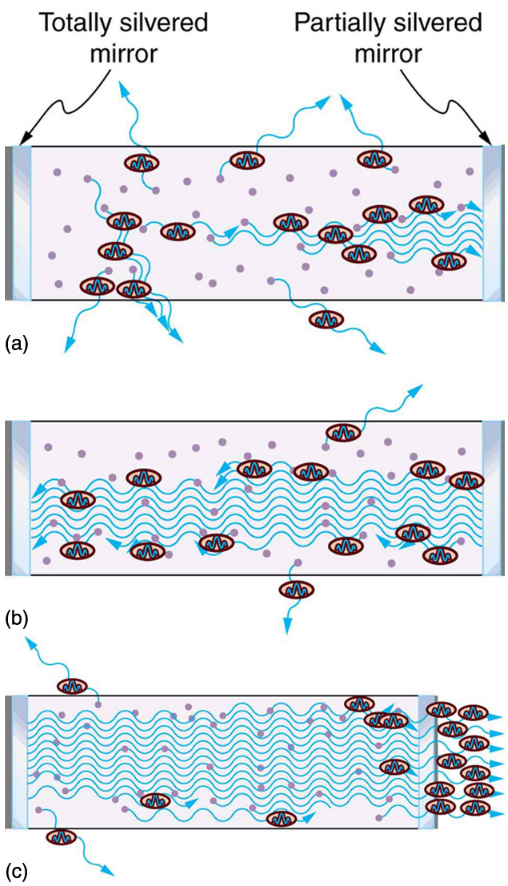
Once a population inversion is achieved, a very interesting thing can happen, as shown in [link]. An electron spontaneously falls from the metastable state, emitting a photon. This photon finds another atom in the metastable state and stimulates it to decay, emitting a second photon of *the same wavelength and in phase* with the first, and so on. **Stimulated emission** is the emission of electromagnetic radiation in the form of photons of a given frequency, triggered by photons of the same frequency. For example, an excited atom, with an electron in an energy orbit higher than normal, releases a photon of a specific frequency when the electron drops back to a lower energy orbit. If this photon then strikes another electron in the same high-energy orbit in another atom, another photon of the same frequency is released. The emitted photons and the triggering photons are always in phase, have the same polarization, and travel in the same direction. The probability of absorption of a photon is the same as the probability of stimulated emission, and so a majority of atoms must be in the metastable state to produce energy. Einstein (again Einstein, and back in 1917!) was one of the important contributors to the understanding of stimulated emission of radiation. Among other things, Einstein was the first to realize that stimulated emission and absorption are equally probable. The laser acts as a temporary energy storage device that subsequently produces a massive energy output of single-wavelength, in-phase photons.



One atom in the metastable state spontaneously decays to a lower level, producing a photon that goes on to stimulate another atom to de-excite. The second photon has exactly the same energy and wavelength as the first and is in phase with it. Both go on to stimulate the emission of other photons. A population inversion is necessary for there to be a net production rather than a net absorption of the photons.

The name **laser** is an acronym for light amplification by stimulated emission of radiation, the process just described. The process was proposed and developed following the advances in quantum physics. A joint Nobel Prize was awarded in 1964 to American Charles Townes (1915–), and Nikolay Basov (1922–2001) and Aleksandr Prokhorov (1916–2002), from the Soviet Union, for the development of lasers. The Nobel Prize in 1981 went to Arthur Schawlow (1921-1999) for pioneering laser applications. The original devices were called masers, because they produced microwaves. The first working laser was created in 1960 at Hughes Research labs (CA) by T. Maiman. It used a pulsed high-powered flash lamp and a ruby rod to produce red light. Today the name laser is used for all such devices developed to produce a variety of wavelengths, including microwave, infrared, visible, and ultraviolet radiation. [link] shows how a laser can be constructed to enhance the stimulated emission of radiation. Energy input can be from a flash tube, electrical discharge, or other sources, in a process sometimes called optical pumping. A large percentage of the original pumping energy is dissipated in other forms, but a population inversion must be achieved. Mirrors can be used to enhance stimulated

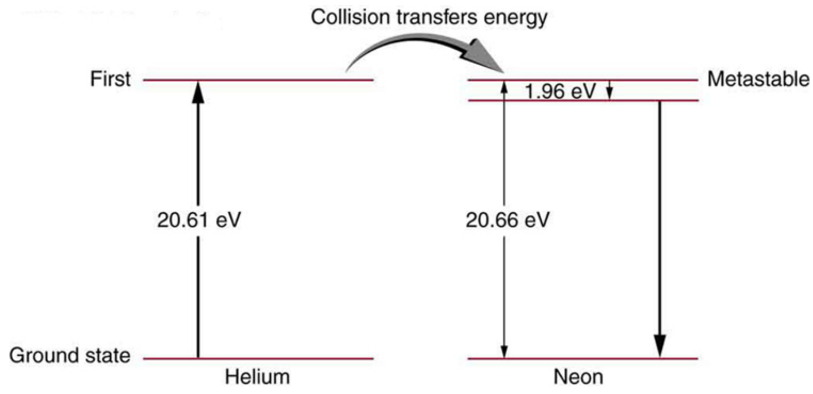
emission by multiple passes of the radiation back and forth through the lasing material. One of the mirrors is semitransparent to allow some of the light to pass through. The laser output from a laser is a mere 1% of the light passing back and forth in a laser.



Typical laser construction has a method of pumping energy into the lasing material to produce a population inversion. (a) Spontaneous emission begins with some photons escaping and others stimulating further emissions. (b) and (c)

Mirrors are used to enhance the probability of stimulated emission by passing photons through the material several times.

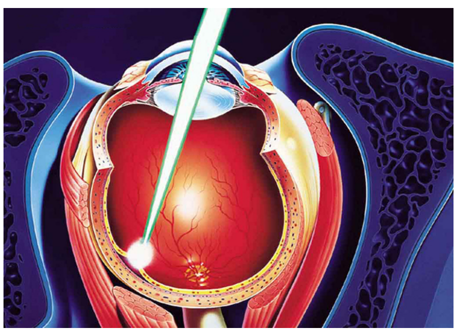
Lasers are constructed from many types of lasing materials, including gases, liquids, solids, and semiconductors. But all lasers are based on the existence of a metastable state or a phosphorescent material. Some lasers produce continuous output; others are pulsed in bursts as brief as 10^{-14} s. Some laser outputs are fantastically powerful—some greater than 10^{12} W—but the more common, everyday lasers produce something on the order of 10^{-3} W. The helium-neon laser that produces a familiar red light is very common. [link] shows the energy levels of helium and neon, a pair of noble gases that work well together. An electrical discharge is passed through a helium-neon gas mixture in which the number of atoms of helium is ten times that of neon. The first excited state of helium is metastable and, thus, stores energy. This energy is easily transferred by collision to neon atoms, because they have an excited state at nearly the same energy as that in helium. That state in neon is also metastable, and this is the one that produces the laser output. (The most likely transition is to the nearby state, producing 1.96 eV photons, which have a wavelength of 633 nm and appear red.) A population inversion can be produced in neon, because there are so many more helium atoms and these put energy into the neon. Helium-neon lasers often have continuous output, because the population inversion can be maintained even while lasing occurs. Probably the most common lasers in use today, including the common laser pointer, are semiconductor or diode lasers, made of silicon. Here, energy is pumped into the material by passing a current in the device to excite the electrons. Special coatings on the ends and fine cleavings of the semiconductor material allow light to bounce back and forth and a tiny fraction to emerge as laser light. Diode lasers can usually run continually and produce outputs in the milliwatt range.



Energy levels in helium and neon. In the common helium-neon laser, an electrical discharge pumps energy into the metastable states of both atoms. The gas mixture has about ten times more helium atoms than neon atoms. Excited helium atoms easily de-excite by transferring energy to neon in a collision. A population inversion in neon is achieved, allowing lasing by the neon to occur.

There are many medical applications of lasers. Lasers have the advantage that they can be focused to a small spot. They also have a well-defined wavelength. Many types of lasers are available today that provide wavelengths from the ultraviolet to the infrared. This is important, as one needs to be able to select a wavelength that will be preferentially absorbed by the material of interest. Objects appear a certain color because they absorb all other visible colors incident upon them. What wavelengths are absorbed depends upon the energy spacing between electron orbitals in that molecule. Unlike the hydrogen atom, biological molecules are complex and have a variety of absorption wavelengths or lines. But these can be determined and used in the selection of a laser with the appropriate wavelength. Water is transparent to the visible spectrum but will absorb light in the UV and IR regions. Blood (hemoglobin) strongly reflects red but absorbs most strongly in the UV.

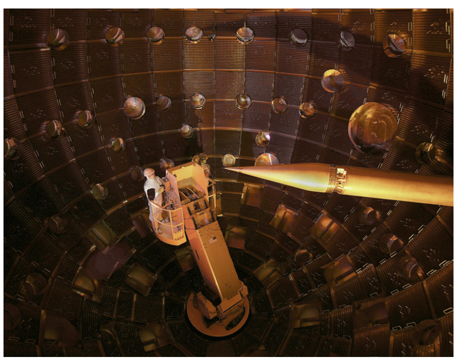
Laser surgery uses a wavelength that is strongly absorbed by the tissue it is focused upon. One example of a medical application of lasers is shown in [link]. A detached retina can result in total loss of vision. Burns made by a laser focused to a small spot on the retina form scar tissue that can hold the retina in place, salvaging the patient’s vision. Other light sources cannot be focused as precisely as a laser due to refractive dispersion of different wavelengths. Similarly, laser surgery in the form of cutting or burning away tissue is made more accurate because laser output can be very precisely focused and is preferentially absorbed because of its single wavelength. Depending upon what part or layer of the retina needs repairing, the appropriate type of laser can be selected. For the repair of tears in the retina, a green argon laser is generally used. This light is absorbed well by tissues containing blood, so coagulation or “welding” of the tear can be done.



A detached retina is burned by a laser designed to focus on a small spot on the retina, the resulting scar tissue holding it in place. The lens of the eye is used to focus the light, as is the device bringing the laser output to the eye.

In dentistry, the use of lasers is rising. Lasers are most commonly used for surgery on the soft tissue of the mouth. They can be used to remove ulcers, stop bleeding, and reshape gum tissue. Their use in cutting into bones and teeth is not quite so common; here the erbium YAG (yttrium aluminum garnet) laser is used.

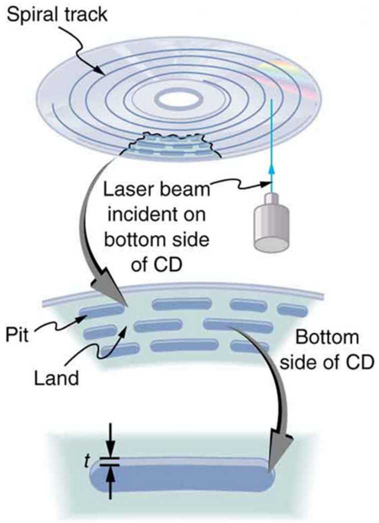
The massive combination of lasers shown in [link] can be used to induce nuclear fusion, the energy source of the sun and hydrogen bombs. Since lasers can produce very high power in very brief pulses, they can be used to focus an enormous amount of energy on a small glass sphere containing fusion fuel. Not only does the incident energy increase the fuel temperature significantly so that fusion can occur, it also compresses the fuel to great density, enhancing the probability of fusion. The compression or implosion is caused by the momentum of the impinging laser photons.



This system of lasers at Lawrence Livermore Laboratory is used to ignite nuclear fusion. A tremendous burst of energy is focused on a small fuel pellet, which is imploded to the high density and temperature needed to make the fusion

reaction proceed. (credit: Lawrence Livermore National Laboratory, Lawrence Livermore National Security, LLC, and the Department of Energy)

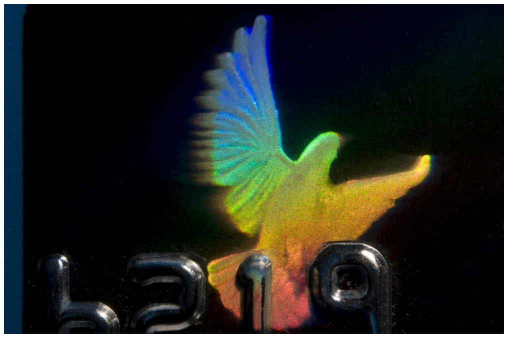
Music CDs are now so common that vinyl records are quaint antiquities. CDs (and DVDs) store information digitally and have a much larger information-storage capacity than vinyl records. An entire encyclopedia can be stored on a single CD. [link] illustrates how the information is stored and read from the CD. Pits made in the CD by a laser can be tiny and very accurately spaced to record digital information. These are read by having an inexpensive solid-state infrared laser beam scatter from pits as the CD spins, revealing their digital pattern and the information encoded upon them.



A CD has digital information

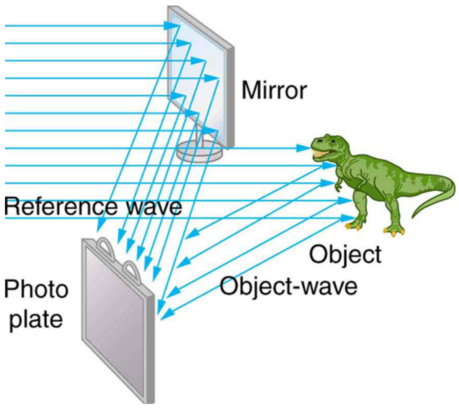
stored in the form of laser-created pits on its surface. These in turn can be read by detecting the laser light scattered from the pit. Large information capacity is possible because of the precision of the laser. Shorter-wavelength lasers enable greater storage capacity.

Holograms, such as those in [link], are true three-dimensional images recorded on film by lasers. Holograms are used for amusement, decoration on novelty items and magazine covers, security on credit cards and driver’s licenses (a laser and other equipment is needed to reproduce them), and for serious three-dimensional information storage. You can see that a hologram is a true three-dimensional image, because objects change relative position in the image when viewed from different angles.



Credit cards commonly have holograms for logos, making them difficult to reproduce (credit: Dominic Alves, Flickr)

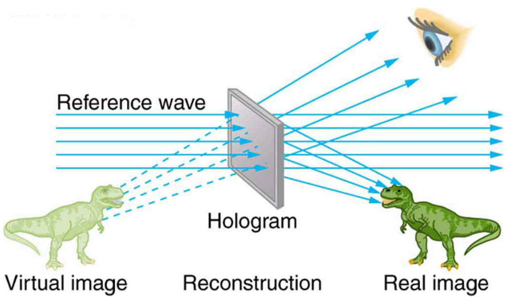
The name **hologram** means “entire picture” (from the Greek *holo*, as in holistic), because the image is three-dimensional. **Holography** is the process of producing holograms and, although they are recorded on photographic film, the process is quite different from normal photography. Holography uses light interference or wave optics, whereas normal photography uses geometric optics. [link] shows one method of producing a hologram. Coherent light from a laser is split by a mirror, with part of the light illuminating the object. The remainder, called the reference beam, shines directly on a piece of film. Light scattered from the object interferes with the reference beam, producing constructive and destructive interference. As a result, the exposed film looks foggy, but close examination reveals a complicated interference pattern stored on it. Where the interference was constructive, the film (a negative actually) is darkened. Holography is sometimes called lensless photography, because it uses the wave characteristics of light as contrasted to normal photography, which uses geometric optics and so requires lenses.



Production of a hologram. Single-wavelength coherent light from a laser produces a well-defined interference pattern on a piece of film. The laser beam is split by a partially silvered mirror, with part of the light illuminating the object and the remainder shining directly on the film.

Light falling on a hologram can form a three-dimensional image. The process is complicated in detail, but the basics can be understood as shown in [link], in which a laser of the same type that exposed the film is now used to illuminate it. The myriad tiny exposed regions of the film are dark and block the light, while less exposed regions allow light to pass. The film thus acts much like a collection of diffraction gratings with various spacings. Light passing through the hologram is diffracted in various directions, producing both real and virtual images of the object used to expose the film. The interference pattern is the same as that produced by the object. Moving

your eye to various places in the interference pattern gives you different perspectives, just as looking directly at the object would. The image thus looks like the object and is three-dimensional like the object.



A transmission hologram is one that produces real and virtual images when a laser of the same type as that which exposed the hologram is passed through it. Diffraction from various parts of the film produces the same interference pattern as the object that was used to expose it.

The hologram illustrated in [link] is a transmission hologram. Holograms that are viewed with reflected light, such as the white light holograms on credit cards, are reflection holograms and are more common. White light holograms often appear a little blurry with rainbow edges, because the diffraction patterns of various colors of light are at slightly different locations due to their different wavelengths. Further uses of holography include all types of 3-D information storage, such as of statues in museums and engineering studies of structures and 3-D images of human organs. Invented in the late 1940s by Dennis Gabor (1900–1970), who won the

1971 Nobel Prize in Physics for his work, holography became far more practical with the development of the laser. Since lasers produce coherent single-wavelength light, their interference patterns are more pronounced. The precision is so great that it is even possible to record numerous holograms on a single piece of film by just changing the angle of the film for each successive image. This is how the holograms that move as you walk by them are produced—a kind of lensless movie.

In a similar way, in the medical field, holograms have allowed complete 3-D holographic displays of objects from a stack of images. Storing these images for future use is relatively easy. With the use of an endoscope, high-resolution 3-D holographic images of internal organs and tissues can be made.

Section Summary

- An important atomic process is fluorescence, defined to be any process in which an atom or molecule is excited by absorbing a photon of a given energy and de-excited by emitting a photon of a lower energy.
- Some states live much longer than others and are termed metastable.
- Phosphorescence is the de-excitation of a metastable state.
- Lasers produce coherent single-wavelength EM radiation by stimulated emission, in which a metastable state is stimulated to decay.
- Lasing requires a population inversion, in which a majority of the atoms or molecules are in their metastable state.

Conceptual Questions

Exercise:

Problem:

How do the allowed orbits for electrons in atoms differ from the allowed orbits for planets around the sun? Explain how the correspondence principle applies here.

Exercise:

Problem:

Atomic and molecular spectra are discrete. What does discrete mean, and how are discrete spectra related to the quantization of energy and electron orbits in atoms and molecules?

Exercise:

Problem:

Hydrogen gas can only absorb EM radiation that has an energy corresponding to a transition in the atom, just as it can only emit these discrete energies. When a spectrum is taken of the solar corona, in which a broad range of EM wavelengths are passed through very hot hydrogen gas, the absorption spectrum shows all the features of the emission spectrum. But when such EM radiation passes through room-temperature hydrogen gas, only the Lyman series is absorbed. Explain the difference.

Exercise:

Problem:

Lasers are used to burn and read CDs. Explain why a laser that emits blue light would be capable of burning and reading more information than one that emits infrared.

Exercise:

Problem:

The coating on the inside of fluorescent light tubes absorbs ultraviolet light and subsequently emits visible light. An inventor claims that he is able to do the reverse process. Is the inventor's claim possible?

Exercise:

Problem:

What is the difference between fluorescence and phosphorescence?

Exercise:

Problem:

How can you tell that a hologram is a true three-dimensional image and that those in 3-D movies are not?

Problem Exercises

Exercise:

Problem:

[link] shows the energy-level diagram for neon. (a) Verify that the energy of the photon emitted when neon goes from its metastable state to the one immediately below is equal to 1.96 eV. (b) Show that the wavelength of this radiation is 633 nm. (c) What wavelength is emitted when the neon makes a direct transition to its ground state?

Solution:

(a) 1.96 eV

(b) $(1240\ \text{eV}\cdot\text{nm})/(1.96\ \text{eV}) = 633\ \text{nm}$

(c) 60.0 nm

Exercise:

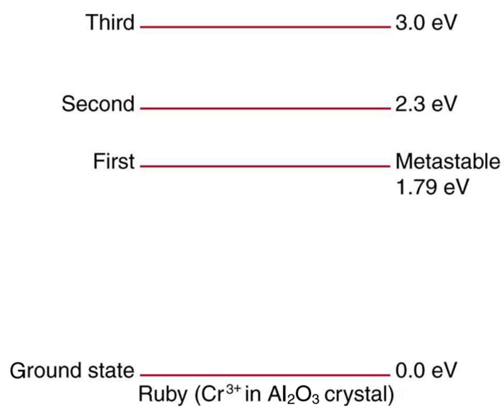
Problem:

A helium-neon laser is pumped by electric discharge. What wavelength electromagnetic radiation would be needed to pump it? See [link] for energy-level information.

Exercise:

Problem:

Ruby lasers have chromium atoms doped in an aluminum oxide crystal. The energy level diagram for chromium in a ruby is shown in [link]. What wavelength is emitted by a ruby laser?



Chromium atoms in an aluminum oxide crystal have these energy levels, one of which is metastable. This is the basis of a ruby laser. Visible light can pump the atom into an excited state above the metastable state to achieve a population inversion.

Solution:

693 nm

Exercise:

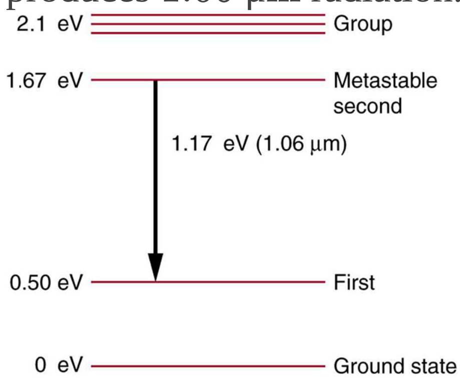
Problem:

(a) What energy photons can pump chromium atoms in a ruby laser from the ground state to its second and third excited states? (b) What are the wavelengths of these photons? Verify that they are in the visible part of the spectrum.

Exercise:

Problem:

Some of the most powerful lasers are based on the energy levels of neodymium in solids, such as glass, as shown in [link]. (a) What average wavelength light can pump the neodymium into the levels above its metastable state? (b) Verify that the 1.17 eV transition produces 1.06 μm radiation.



Neodymium atoms in glass have these energy levels, one of which is metastable. The group of levels above the metastable state is convenient for achieving a population inversion, since photons of many different energies can be absorbed by atoms in the ground state.

Solution:

(a) 590 nm

(b) $(1240 \text{ eV}\cdot\text{nm})/(1.17 \text{ eV}) = 1.06 \ \mu\text{m}$

Glossary

metastable
: a state whose lifetime is an order of magnitude longer than the most short-lived states

atomic excitation
: a state in which an atom or ion acquires the necessary energy to promote one or more of its electrons to electronic states higher in energy than their ground state

atomic de-excitation
: process by which an atom transfers from an excited electronic state back to the ground state electronic configuration; often occurs by emission of a photon

laser
: acronym for light amplification by stimulated emission of radiation

phosphorescence
: the de-excitation of a metastable state

population inversion
: the condition in which the majority of atoms in a sample are in a metastable state

stimulated emission
: emission by atom or molecule in which an excited state is stimulated to decay, most readily caused by a photon of the same energy that is necessary to excite the state

hologram
: means *entire picture* (from the Greek word *holo*, as in holistic), because the image produced is three dimensional

holography
: the process of producing holograms

fluorescence

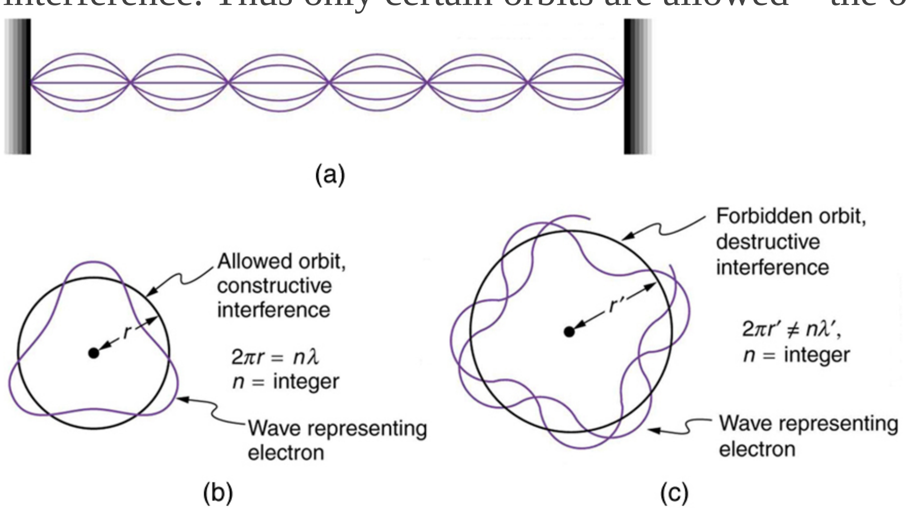
any process in which an atom or molecule, excited by a photon of a given energy, de-excites by emission of a lower-energy photon

The Wave Nature of Matter Causes Quantization

- Explain Bohr's model of atom.
- Define and describe quantization of angular momentum.
- Calculate the angular momentum for an orbit of atom.
- Define and describe the wave-like properties of matter.

After visiting some of the applications of different aspects of atomic physics, we now return to the basic theory that was built upon Bohr's atom. Einstein once said it was important to keep asking the questions we eventually teach children not to ask. Why is angular momentum quantized? You already know the answer. Electrons have wave-like properties, as de Broglie later proposed. They can exist only where they interfere constructively, and only certain orbits meet proper conditions, as we shall see in the next module.

Following Bohr's initial work on the hydrogen atom, a decade was to pass before de Broglie proposed that matter has wave properties. The wave-like properties of matter were subsequently confirmed by observations of electron interference when scattered from crystals. Electrons can exist only in locations where they interfere constructively. How does this affect electrons in atomic orbits? When an electron is bound to an atom, its wavelength must fit into a small space, something like a standing wave on a string. (See [link].) Allowed orbits are those orbits in which an electron constructively interferes with itself. Not all orbits produce constructive interference. Thus only certain orbits are allowed—the orbits are quantized.



(a) Waves on a string have a wavelength related to the length of the string, allowing them to interfere constructively. (b) If we imagine the string bent into a closed circle, we get a rough idea of how electrons in circular orbits can interfere constructively. (c) If the wavelength does not fit into the circumference, the electron interferes destructively; it cannot exist in such an orbit.

For a circular orbit, constructive interference occurs when the electron's wavelength fits neatly into the circumference, so that wave crests always align with crests and wave troughs align with troughs, as shown in [link] (b). More precisely, when an integral multiple of the electron's wavelength equals the circumference of the orbit, constructive interference is obtained. In equation form, the *condition for constructive interference and an allowed electron orbit* is

Equation:

$$n\lambda_n = 2\pi r_n (n = 1, 2, 3 \ldots),$$

where λ_n is the electron's wavelength and r_n is the radius of that circular orbit. The de Broglie wavelength is $\lambda = h/p = h/\ mv$, and so here $\lambda = h/m_e v$. Substituting this into the previous condition for constructive interference produces an interesting result:

Equation:

$$\frac{\text{nh}}{m_e v} = 2\text{π}r_n.$$

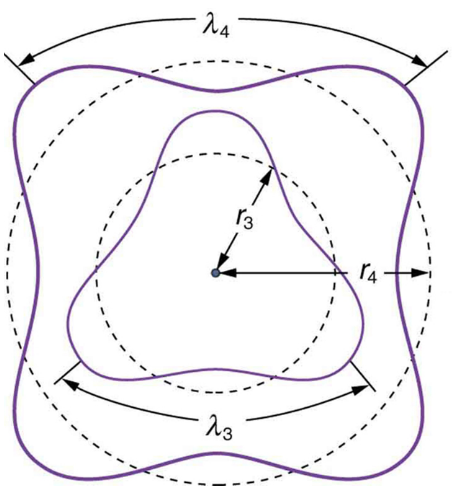
Rearranging terms, and noting that $L = \text{mvr}$ for a circular orbit, we obtain the quantization of angular momentum as the condition for allowed orbits:

Equation:

$$L = m_e \mathrm{vr}_n = n\frac{h}{2\pi}(n = 1, 2, 3 \ldots).$$

This is what Bohr was forced to hypothesize as the rule for allowed orbits, as stated earlier. We now realize that it is the condition for constructive interference of an electron in a circular orbit. [link] illustrates this for $n = 3$ and $n = 4$.

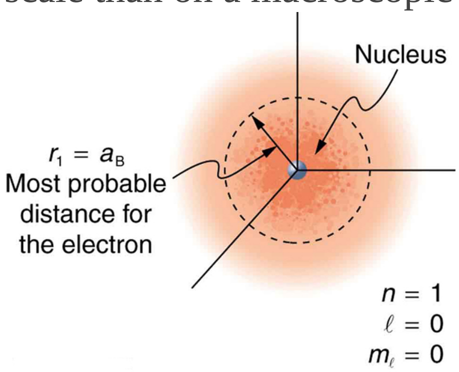
Note:
Waves and Quantization
The wave nature of matter is responsible for the quantization of energy levels in bound systems. Only those states where matter interferes constructively exist, or are "allowed." Since there is a lowest orbit where this is possible in an atom, the electron cannot spiral into the nucleus. It cannot exist closer to or inside the nucleus. The wave nature of matter is what prevents matter from collapsing and gives atoms their sizes.



The third and fourth allowed circular orbits have three and four wavelengths,

respectively, in their circumferences.

Because of the wave character of matter, the idea of well-defined orbits gives way to a model in which there is a cloud of probability, consistent with Heisenberg's uncertainty principle. [link] shows how this applies to the ground state of hydrogen. If you try to follow the electron in some well-defined orbit using a probe that has a small enough wavelength to get some details, you will instead knock the electron out of its orbit. Each measurement of the electron's position will find it to be in a definite location somewhere near the nucleus. Repeated measurements reveal a cloud of probability like that in the figure, with each speck the location determined by a single measurement. There is not a well-defined, circular-orbit type of distribution. Nature again proves to be different on a small scale than on a macroscopic scale.



The ground state of a hydrogen atom has a probability cloud describing the position of its electron. The probability of finding the electron is proportional to the darkness of the cloud. The electron can be closer or farther than

the Bohr radius, but it is very unlikely to be a great distance from the nucleus.

There are many examples in which the wave nature of matter causes quantization in bound systems such as the atom. Whenever a particle is confined or bound to a small space, its allowed wavelengths are those which fit into that space. For example, the particle in a box model describes a particle free to move in a small space surrounded by impenetrable barriers. This is true in blackbody radiators (atoms and molecules) as well as in atomic and molecular spectra. Various atoms and molecules will have different sets of electron orbits, depending on the size and complexity of the system. When a system is large, such as a grain of sand, the tiny particle waves in it can fit in so many ways that it becomes impossible to see that the allowed states are discrete. Thus the correspondence principle is satisfied. As systems become large, they gradually look less grainy, and quantization becomes less evident. Unbound systems (small or not), such as an electron freed from an atom, do not have quantized energies, since their wavelengths are not constrained to fit in a certain volume.

Note:
PhET Explorations: Quantum Wave Interference
When do photons, electrons, and atoms behave like particles and when do they behave like waves? Watch waves spread out and interfere as they pass through a double slit, then get detected on a screen as tiny dots. Use quantum detectors to explore how measurements change the waves and the patterns they produce on the screen.

Quantum Wave

Interference

Section Summary

- Quantization of orbital energy is caused by the wave nature of matter. Allowed orbits in atoms occur for constructive interference of electrons in the orbit, requiring an integral number of wavelengths to fit in an orbit's circumference; that is,
  Equation:
  $$n\lambda_n = 2\pi r_n (n = 1, 2, 3 \ldots),$$
  where λ_n is the electron's de Broglie wavelength.
- Owing to the wave nature of electrons and the Heisenberg uncertainty principle, there are no well-defined orbits; rather, there are clouds of probability.
- Bohr correctly proposed that the energy and radii of the orbits of electrons in atoms are quantized, with energy for transitions between orbits given by
  Equation:
  $$\Delta E = hf = E_\mathrm{i} - E_\mathrm{f},$$
  where ΔE is the change in energy between the initial and final orbits and hf is the energy of an absorbed or emitted photon.
- It is useful to plot orbit energies on a vertical graph called an energy-level diagram.
- The allowed orbits are circular, Bohr proposed, and must have quantized orbital angular momentum given by
  Equation:
  $$L = m_e v r_n = n\frac{h}{2\pi} (n = 1, 2, 3 \ldots),$$

where L is the angular momentum, r_n is the radius of orbit n, and h is Planck’s constant.

Conceptual Questions

Exercise:

Problem:

How is the de Broglie wavelength of electrons related to the quantization of their orbits in atoms and molecules?

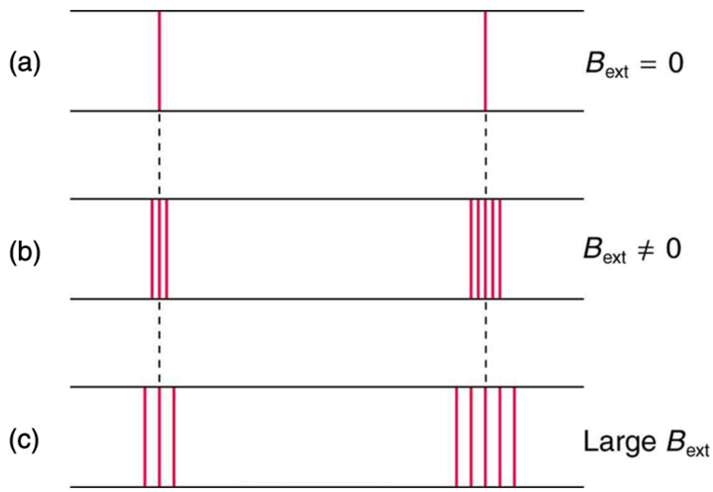
Patterns in Spectra Reveal More Quantization

- State and discuss the Zeeman effect.
- Define orbital magnetic field.
- Define orbital angular momentum.
- Define space quantization.

High-resolution measurements of atomic and molecular spectra show that the spectral lines are even more complex than they first appear. In this section, we will see that this complexity has yielded important new information about electrons and their orbits in atoms.

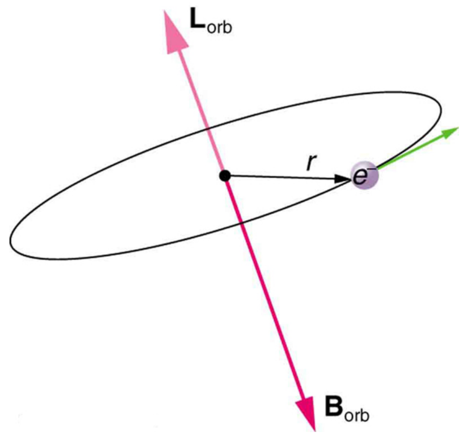
In order to explore the substructure of atoms (and knowing that magnetic fields affect moving charges), the Dutch physicist Hendrik Lorentz (1853–1930) suggested that his student Pieter Zeeman (1865–1943) study how spectra might be affected by magnetic fields. What they found became known as the **Zeeman effect**, which involved spectral lines being split into two or more separate emission lines by an external magnetic field, as shown in [link]. For their discoveries, Zeeman and Lorentz shared the 1902 Nobel Prize in Physics.

Zeeman splitting is complex. Some lines split into three lines, some into five, and so on. But one general feature is that the amount the split lines are separated is proportional to the applied field strength, indicating an interaction with a moving charge. The splitting means that the quantized energy of an orbit is affected by an external magnetic field, causing the orbit to have several discrete energies instead of one. Even without an external magnetic field, very precise measurements showed that spectral lines are doublets (split into two), apparently by magnetic fields within the atom itself.

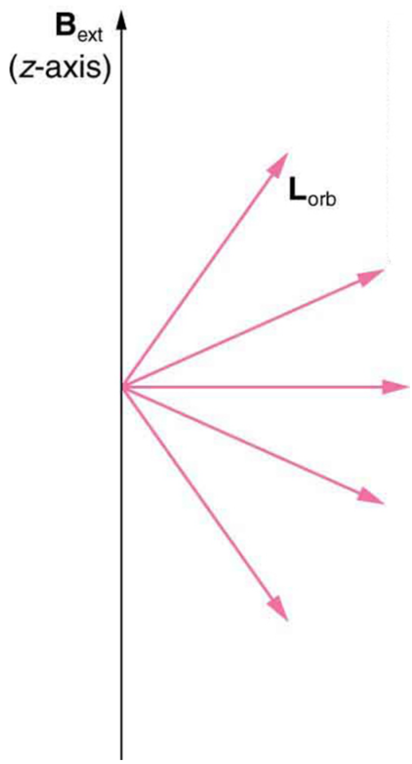


The Zeeman effect is the splitting of spectral lines when a magnetic field is applied. The number of lines formed varies, but the spread is proportional to the strength of the applied field. (a) Two spectral lines with no external magnetic field. (b) The lines split when the field is applied. (c) The splitting is greater when a stronger field is applied.

Bohr's theory of circular orbits is useful for visualizing how an electron's orbit is affected by a magnetic field. The circular orbit forms a current loop, which creates a magnetic field of its own, $\mathbf{B}_{orb}$ as seen in [link]. Note that the **orbital magnetic field** $\mathbf{B}_{orb}$ and the **orbital angular momentum** $\mathbf{L}_{orb}$ are along the same line. The external magnetic field and the orbital magnetic field interact; a torque is exerted to align them. A torque rotating a system through some angle does work so that there is energy associated with this interaction. Thus, orbits at different angles to the external magnetic field have different energies. What is remarkable is that the energies are quantized—the magnetic field splits the spectral lines into several discrete lines that have different energies. This means that only certain angles are allowed between the orbital angular momentum and the external field, as seen in [link].



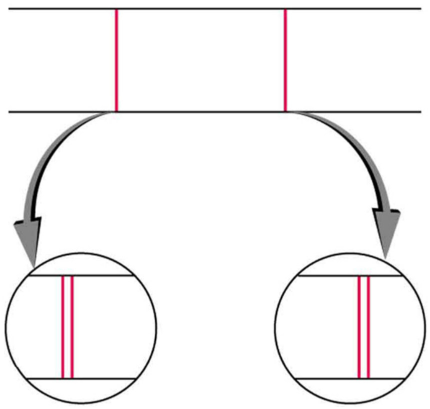
The approximate picture of an electron in a circular orbit illustrates how the current loop produces its own magnetic field, called $\mathbf{B}_{\text{orb}}$. It also shows how $\mathbf{B}_{\text{orb}}$ is along the same line as the orbital angular momentum $\mathbf{L}_{\text{orb}}$.



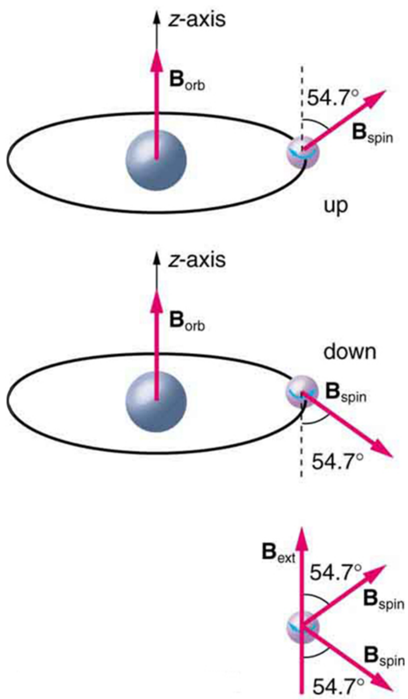
Only certain angles are allowed between the orbital angular momentum and an external magnetic field. This is implied by the fact that the Zeeman effect splits spectral lines into several discrete lines. Each line is associated with an angle between the external magnetic field and magnetic fields due to electrons and their orbits.

We already know that the magnitude of angular momentum is quantized for electron orbits in atoms. The new insight is that the *direction of the orbital angular momentum is also quantized*. The fact that the orbital angular momentum can have only certain directions is called **space quantization**. Like many aspects of quantum mechanics, this quantization of direction is totally unexpected. On the macroscopic scale, orbital angular momentum, such as that of the moon around the earth, can have any magnitude and be in any direction.

Detailed treatment of space quantization began to explain some complexities of atomic spectra, but certain patterns seemed to be caused by something else. As mentioned, spectral lines are actually closely spaced doublets, a characteristic called **fine structure**, as shown in [link]. The doublet changes when a magnetic field is applied, implying that whatever causes the doublet interacts with a magnetic field. In 1925, Sem Goudsmit and George Uhlenbeck, two Dutch physicists, successfully argued that electrons have properties analogous to a macroscopic charge spinning on its axis. Electrons, in fact, have an internal or intrinsic angular momentum called **intrinsic spin $\mathbf{S}$**. Since electrons are charged, their intrinsic spin creates an **intrinsic magnetic field $\mathbf{B}_{\text{int}}$**, which interacts with their orbital magnetic field $\mathbf{B}_{\text{orb}}$. Furthermore, *electron intrinsic spin is quantized in magnitude and direction*, analogous to the situation for orbital angular momentum. The spin of the electron can have only one magnitude, and its direction can be at only one of two angles relative to a magnetic field, as seen in [link]. We refer to this as spin up or spin down for the electron. Each spin direction has a different energy; hence, spectroscopic lines are split into two. Spectral doublets are now understood as being due to electron spin.



Fine structure. Upon close examination, spectral lines are doublets, even in the absence of an external magnetic field. The electron has an intrinsic magnetic field that interacts with its orbital magnetic field.



The intrinsic magnetic field $\mathbf{B}_{\text{int}}$ of an electron is attributed to its spin, $\mathbf{S}$, roughly pictured to be due to its charge spinning on its axis. This is only a crude model, since electrons seem to have no size. The spin and intrinsic magnetic field of the electron can make only one of two angles with another magnetic field, such as that created by the electron's orbital

motion. Space is quantized for spin as well as for orbital angular momentum.

These two new insights—that the direction of angular momentum, whether orbital or spin, is quantized, and that electrons have intrinsic spin—help to explain many of the complexities of atomic and molecular spectra. In magnetic resonance imaging, it is the way that the intrinsic magnetic field of hydrogen and biological atoms interact with an external field that underlies the diagnostic fundamentals.

Section Summary

- The Zeeman effect—the splitting of lines when a magnetic field is applied—is caused by other quantized entities in atoms.
- Both the magnitude and direction of orbital angular momentum are quantized.
- The same is true for the magnitude and direction of the intrinsic spin of electrons.

Conceptual Questions

Exercise:

Problem:

What is the Zeeman effect, and what type of quantization was discovered because of this effect?

Glossary

Zeeman effect

the effect of external magnetic fields on spectral lines

intrinsic spin
: the internal or intrinsic angular momentum of electrons

orbital angular momentum
: an angular momentum that corresponds to the quantum analog of classical angular momentum

fine structure
: the splitting of spectral lines of the hydrogen spectrum when the spectral lines are examined at very high resolution

space quantization
: the fact that the orbital angular momentum can have only certain directions

intrinsic magnetic field
: the magnetic field generated due to the intrinsic spin of electrons

orbital magnetic field
: the magnetic field generated due to the orbital motion of electrons

Quantum Numbers and Rules

- Define quantum number.
- Calculate angle of angular momentum vector with an axis.
- Define spin quantum number.

Physical characteristics that are quantized—such as energy, charge, and angular momentum—are of such importance that names and symbols are given to them. The values of quantized entities are expressed in terms of **quantum numbers**, and the rules governing them are of the utmost importance in determining what nature is and does. This section covers some of the more important quantum numbers and rules—all of which apply in chemistry, material science, and far beyond the realm of atomic physics, where they were first discovered. Once again, we see how physics makes discoveries which enable other fields to grow.

The *energy states of bound systems are quantized*, because the particle wavelength can fit into the bounds of the system in only certain ways. This was elaborated for the hydrogen atom, for which the allowed energies are expressed as $E_n \propto 1/n^2$, where $n = 1, 2, 3, ...$ We define n to be the principal quantum number that labels the basic states of a system. The lowest-energy state has $n = 1$, the first excited state has $n = 2$, and so on. Thus the allowed values for the principal quantum number are

Equation:

$$n = 1, 2, 3, ...$$

This is more than just a numbering scheme, since the energy of the system, such as the hydrogen atom, can be expressed as some function of n, as can other characteristics (such as the orbital radii of the hydrogen atom).

The fact that the *magnitude of angular momentum is quantized* was first recognized by Bohr in relation to the hydrogen atom; it is now known to be true in general. With the development of quantum mechanics, it was found that the magnitude of angular momentum L can have only the values

Equation:

$$L = \sqrt{l(l+1)}\frac{h}{2\pi} \quad (l = 0, 1, 2, ..., n-1),$$

where l is defined to be the **angular momentum quantum number**. The rule for l in atoms is given in the parentheses. Given n, the value of l can be any integer from zero up to $n-1$. For example, if $n = 4$, then l can be 0, 1, 2, or 3.

Note that for $n = 1$, l can only be zero. This means that the ground-state angular momentum for hydrogen is actually zero, not $h/2\pi$ as Bohr proposed. The picture of circular orbits is not valid, because there would be angular momentum for any circular orbit. A more valid picture is the cloud of probability shown for the ground state of hydrogen in [link]. The electron actually spends time in and near the nucleus. The reason the electron does not remain in the nucleus is related to Heisenberg's uncertainty principle—the electron's energy would have to be much too large to be confined to the small space of the nucleus. Now the first excited state of hydrogen has $n = 2$, so that l can be either 0 or 1, according to the rule in $L = \sqrt{l(l+1)}\frac{h}{2\pi}$. Similarly, for $n = 3$, l can be 0, 1, or 2. It is often most convenient to state the value of l, a simple integer, rather than calculating the value of L from $L = \sqrt{l(l+1)}\frac{h}{2\pi}$. For example, for $l = 2$, we see that

Equation:

$$L = \sqrt{2(2+1)}\frac{h}{2\pi} = \sqrt{6}\frac{h}{2\pi} = 0.390h = 2.58 \times 10^{-34}\ \text{J} \cdot \text{s}.$$

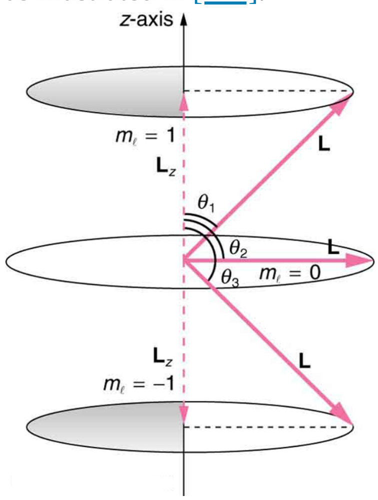
It is much simpler to state $l = 2$.

As recognized in the Zeeman effect, the *direction of angular momentum is quantized*. We now know this is true in all circumstances. It is found that the component of angular momentum along one direction in space, usually called the z-axis, can have only certain values of L_z. The direction in space must be related to something physical, such as the direction of the magnetic field at that location. This is an aspect of relativity. Direction has no meaning if there is nothing that varies with direction, as does magnetic force. The allowed values of L_z are

Equation:

$$L_z = m_l\frac{h}{2\pi} \quad (m_l = -l, -l+1, ..., \ -1, 0, 1, ... \ l-1, l),$$

where L_z is the **z-component of the angular momentum** and m_l is the angular momentum projection quantum number. The rule in parentheses for the values of m_l is that it can range from $-l$ to l in steps of one. For example, if $l = 2$, then m_l can have the five values –2, –1, 0, 1, and 2. Each m_l corresponds to a different energy in the presence of a magnetic field, so that they are related to the splitting of spectral lines into discrete parts, as discussed in the preceding section. If the z-component of angular momentum can have only certain values, then the angular momentum can have only certain directions, as illustrated in [link].



The component of a given angular momentum along the z-axis (defined by the direction of a magnetic field) can have only certain values; these are shown here for $l = 1$, for which $m_l = -1, 0, \text{and} +1$.

The direction of L is quantized in the sense that it can have only certain angles relative to the z-axis.

Example:
What Are the Allowed Directions?

Calculate the angles that the angular momentum vector $\mathbf{L}$ can make with the z-axis for $l = 1$, as illustrated in [link].

Strategy

[link] represents the vectors $\mathbf{L}$ and $\mathbf{L}_z$ as usual, with arrows proportional to their magnitudes and pointing in the correct directions. $\mathbf{L}$ and $\mathbf{L}_z$ form a right triangle, with $\mathbf{L}$ being the hypotenuse and $\mathbf{L}_z$ the adjacent side. This means that the ratio of $\mathbf{L}_z$ to $\mathbf{L}$ is the cosine of the angle of interest. We can find $\mathbf{L}$ and $\mathbf{L}_z$ using $L = \sqrt{l(l+1)}\frac{h}{2\pi}$ and $L_z = m\frac{h}{2\pi}$.

Solution

We are given $l = 1$, so that m_l can be +1, 0, or −1. Thus L has the value given by $L = \sqrt{l(l+1)}\frac{h}{2\pi}$.

Equation:

$$L = \frac{\sqrt{l(l+1)}h}{2\pi} = \frac{\sqrt{2}h}{2\pi}$$

L_z can have three values, given by $L_z = m_l\frac{h}{2\pi}$.

Equation:

$$L_z = m_l\frac{h}{2\pi} = \begin{array}{lll} \frac{h}{2\pi}, & m_l & = & +1 \\ 0, & m_l & = & 0 \\ -\frac{h}{2\pi}, & m_l & = & -1 \end{array}$$

As can be seen in [link], $\cos\theta = \mathrm{L}_z/\mathrm{L}$, and so for $m_l = +1$, we have

Equation:

$$\cos\theta_1 = \frac{L_Z}{L} = \frac{\frac{h}{2\pi}}{\frac{\sqrt{2}h}{2\pi}} = \frac{1}{\sqrt{2}} = 0.707.$$

Thus,

Equation:

$$\theta_1 = \cos^{-1}0.707 = 45.0^\circ.$$

Similarly, for $m_l = 0$, we find $\cos\theta_2 = 0$; thus,

Equation:

$$\theta_2 = \cos^{-1} 0 = 90.0^{\circ}.$$

And for $m_l = -1$,
Equation:

$$\cos\theta_3 = \frac{L_Z}{L} = \frac{-\frac{h}{2\pi}}{\frac{\sqrt{2}h}{2\pi}} = -\frac{1}{\sqrt{2}} = -0.707,$$

so that
Equation:

$$\theta_3 = \cos^{-1}(-0.707) = 135.0^{\circ}.$$

Discussion
The angles are consistent with the figure. Only the angle relative to the z-axis is quantized. L can point in any direction as long as it makes the proper angle with the z-axis. Thus the angular momentum vectors lie on cones as illustrated. This behavior is not observed on the large scale. To see how the correspondence principle holds here, consider that the smallest angle (θ_1 in the example) is for the maximum value of $m_l = 0$, namely $m_l = l$. For that smallest angle,
Equation:

$$\cos\theta = \frac{L_z}{L} = \frac{l}{\sqrt{l(l+1)}},$$

which approaches 1 as l becomes very large. If $\cos\theta = 1$, then $\theta = 0^{\circ}$. Furthermore, for large l, there are many values of m_l, so that all angles become possible as l gets very large.

Intrinsic Spin Angular Momentum Is Quantized in Magnitude and Direction

There are two more quantum numbers of immediate concern. Both were first discovered for electrons in conjunction with fine structure in atomic spectra. It is now well established that electrons and other fundamental particles have *intrinsic spin*, roughly analogous to a planet spinning on its axis. This spin is a fundamental characteristic of particles, and only one magnitude of intrinsic spin is allowed for a given type of particle. Intrinsic angular momentum is quantized independently of orbital angular momentum. Additionally, the direction of the spin is also quantized. It has been found that the **magnitude of the intrinsic (internal) spin angular momentum**, S, of an electron is given by
Equation:

$$S = \sqrt{s(s+1)}\frac{h}{2\pi} \quad (s = 1/2 \text{ for electrons}),$$

where s is defined to be the **spin quantum number**. This is very similar to the quantization of L given in $L = \sqrt{l(l+1)}\frac{h}{2\pi}$, except that the only value allowed for s for electrons is 1/2.

The *direction of intrinsic spin is quantized*, just as is the direction of orbital angular momentum. The direction of spin angular momentum along one direction in space, again called the z-axis, can have only the values

Equation:

$$S_z = m_s \frac{h}{2\pi} \quad \left(m_s = -\frac{1}{2}, +\frac{1}{2}\right)$$

for electrons. S_z is the z**-component of spin angular momentum** and m_s is the **spin projection quantum number**. For electrons, s can only be 1/2, and m_s can be either +1/2 or –1/2. Spin projection m_s=$+1/2$ is referred to as *spin up*, whereas $m_s = -1/2$ is called *spin down*. These are illustrated in [link].

Note:
Intrinsic Spin
In later chapters, we will see that intrinsic spin is a characteristic of all subatomic particles. For some particles s is half-integral, whereas for others s is integral—there are crucial differences between half-integral spin particles and integral spin particles. Protons and neutrons, like electrons, have $s = 1/2$, whereas photons have $s = 1$, and other particles called pions have $s = 0$, and so on.

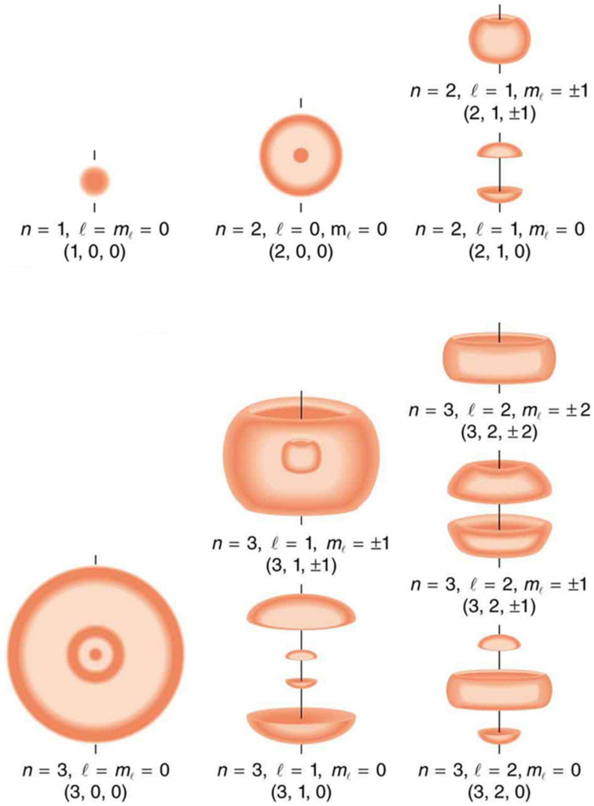
To summarize, the state of a system, such as the precise nature of an electron in an atom, is determined by its particular quantum numbers. These are expressed in the form (n, l, m_l, m_s) —see [link] *For electrons in atoms*, the principal quantum number can have the values $n = 1, 2, 3, \ldots$. Once n is known, the values of the angular momentum quantum number are limited to $l = 1, 2, 3, \ldots, n - 1$. For a given value of l, the angular momentum projection quantum number can have only the values $m_l = -l, \ -l\ +1, \ldots, -1, 0, 1, \ldots, l - 1, l$. Electron spin is independent of n, l, and m_l, always having $s = 1/2$. The spin projection quantum number can have two values, $m_s = 1/2$ or $\ -1/2$.

Name	Symbol	Allowed values
Principal quantum number	n	$1, 2, 3, \ldots$
Angular momentum	l	$0, 1, 2, \ldots n - 1$
Angular momentum projection	m_l	$-l, \ -l\ +1, \ldots, \ -1, 0, 1, \ldots, l-1, l$ (or $0, \pm1, \pm2, \ldots, \ \pm l$)

Name	Symbol	Allowed values
Spin[footnote] The spin quantum number *s* is usually not stated, since it is always 1/2 for electrons	s	$1/2(\text{electrons})$
Spin projection	m_s	$-1/2,\ +1/2$

Atomic Quantum Numbers

[link] shows several hydrogen states corresponding to different sets of quantum numbers. Note that these clouds of probability are the locations of electrons as determined by making repeated measurements—each measurement finds the electron in a definite location, with a greater chance of finding the electron in some places rather than others. With repeated measurements, the pattern of probability shown in the figure emerges. The clouds of probability do not look like nor do they correspond to classical orbits. The uncertainty principle actually prevents us and nature from knowing how the electron gets from one place to another, and so an orbit really does not exist as such. Nature on a small scale is again much different from that on the large scale.



Probability clouds for the electron in the ground state and several excited states of hydrogen. The nature of these states is determined by their sets of quantum numbers, here given as (n, l, m_l). The ground state is (0, 0, 0); one of the possibilities for the second excited state is (3, 2, 1). The probability of finding the electron is indicated by the shade of color; the darker the coloring the greater the chance of finding the electron.

We will see that the quantum numbers discussed in this section are valid for a broad range of particles and other systems, such as nuclei. Some quantum numbers, such as intrinsic spin, are related to fundamental classifications of subatomic particles, and they obey laws that will give us further insight into the substructure of matter and its interactions.

Note:
PhET Explorations: Stern-Gerlach Experiment
The classic Stern-Gerlach Experiment shows that atoms have a property called spin. Spin is a kind of intrinsic angular momentum, which has no classical counterpart. When the z-component of the spin is

measured, one always gets one of two values: spin up or spin down.

https://phet.colorado.edu/sims/stern-gerlach/stern-gerlach_en.html

Section Summary

- Quantum numbers are used to express the allowed values of quantized entities. The principal quantum number n labels the basic states of a system and is given by
  Equation:
  $$n = 1, 2, 3, \ldots .$$
- The magnitude of angular momentum is given by
  Equation:
  $$L = \sqrt{l(l+1)}\frac{h}{2\pi} \quad (l = 0, 1, 2, \ldots, n-1),$$
  where l is the angular momentum quantum number. The direction of angular momentum is quantized, in that its component along an axis defined by a magnetic field, called the z-axis is given by
  Equation:
  $$L_z = m_l\frac{h}{2\pi} \quad (m_l = -l, -l+1, \ldots, -1, 0, 1, \ldots l-1, l),$$
  where L_z is the z-component of the angular momentum and m_l is the angular momentum projection quantum number. Similarly, the electron's intrinsic spin angular momentum S is given by
  Equation:
  $$S = \sqrt{s(s+1)}\frac{h}{2\pi} \quad (s = 1/2 \text{ for electrons}),$$
  s is defined to be the spin quantum number. Finally, the direction of the electron's spin along the z-axis is given by
  Equation:
  $$S_z = m_s\frac{h}{2\pi} \quad \left(m_s = -\frac{1}{2}, +\frac{1}{2}\right),$$
  where S_z is the z-component of spin angular momentum and m_s is the spin projection quantum number. Spin projection m_s=$+1/2$ is referred to as spin up, whereas $m_s = -1/2$ is called spin down. [link] summarizes the atomic quantum numbers and their allowed values.

Conceptual Questions

Exercise:

Problem: Define the quantum numbers n, l, m_l, s, and m_s.

Exercise:

Problem: For a given value of n, what are the allowed values of l?

Exercise:

Problem:

For a given value of l, what are the allowed values of m_l? What are the allowed values of m_l for a given value of n? Give an example in each case.

Exercise:

Problem:

List all the possible values of s and m_s for an electron. Are there particles for which these values are different? The same?

Problem Exercises

Exercise:

Problem:

If an atom has an electron in the $n = 5$ state with $m_l = 3$, what are the possible values of l?

Solution:

$l = 4$, 3 are possible since $l < n$ and $\mid m_l \mid \leq l$.

Exercise:

Problem: An atom has an electron with $m_l = 2$. What is the smallest value of n for this electron?

Exercise:

Problem: What are the possible values of m_l for an electron in the $n = 4$ state?

Solution:

$n = 4 \Rightarrow l = 3, 2, 1, 0 \Rightarrow m_l = \pm 3, \ \pm 2, \ \pm 1, 0$ are possible.

Exercise:

Problem:

What, if any, constraints does a value of $m_l = 1$ place on the other quantum numbers for an electron in an atom?

Exercise:

Problem:

(a) Calculate the magnitude of the angular momentum for an $l=1$ electron. (b) Compare your answer to the value Bohr proposed for the $n=1$ state.

Solution:

(a) $1.49\times10^{-34}\ \mathrm{J\cdot s}$

(b) $1.06\times10^{-34}\ \mathrm{J\cdot s}$

Exercise:

Problem:

(a) What is the magnitude of the angular momentum for an $l=1$ electron? (b) Calculate the magnitude of the electron's spin angular momentum. (c) What is the ratio of these angular momenta?

Exercise:

Problem: Repeat [link] for $l=3$.

Solution:

(a) $3.66\times10^{-34}\ \mathrm{J\cdot s}$

(b) $s=9.13\times10^{-35}\ \mathrm{J\cdot s}$

(c) $\frac{L}{S}=\frac{\sqrt{12}}{\sqrt{3/4}}=4$

Exercise:

Problem:

(a) How many angles can L make with the z-axis for an $l=2$ electron? (b) Calculate the value of the smallest angle.

Exercise:

Problem: What angles can the spin S of an electron make with the z-axis?

Solution:

$\theta=54.7^{\mathrm{o}}$, 125.3^{o}

Glossary

quantum numbers
: the values of quantized entities, such as energy and angular momentum

angular momentum quantum number

a quantum number associated with the angular momentum of electrons

spin quantum number
: the quantum number that parameterizes the intrinsic angular momentum (or spin angular momentum, or simply spin) of a given particle

spin projection quantum number
: quantum number that can be used to calculate the intrinsic electron angular momentum along the z-axis

z-component of spin angular momentum
: component of intrinsic electron spin along the z-axis

magnitude of the intrinsic (internal) spin angular momentum
: given by $S = \sqrt{s(s+1)}\frac{h}{2\pi}$

z-component of the angular momentum
: component of orbital angular momentum of electron along the z-axis

The Pauli Exclusion Principle

- Define the composition of an atom along with its electrons, neutrons, and protons.
- Explain the Pauli exclusion principle and its application to the atom.
- Specify the shell and subshell symbols and their positions.
- Define the position of electrons in different shells of an atom.
- State the position of each element in the periodic table according to shell filling.

Multiple-Electron Atoms

All atoms except hydrogen are multiple-electron atoms. The physical and chemical properties of elements are directly related to the number of electrons a neutral atom has. The periodic table of the elements groups elements with similar properties into columns. This systematic organization is related to the number of electrons in a neutral atom, called the **atomic number**, Z. We shall see in this section that the exclusion principle is key to the underlying explanations, and that it applies far beyond the realm of atomic physics.

In 1925, the Austrian physicist Wolfgang Pauli (see [link]) proposed the following rule: No two electrons can have the same set of quantum numbers. That is, no two electrons can be in the same state. This statement is known as the **Pauli exclusion principle**, because it excludes electrons from being in the same state. The Pauli exclusion principle is extremely powerful and very broadly applicable. It applies to any identical particles with half-integral intrinsic spin—that is, having $s=1/2,\ 3/2,\ \ldots$ Thus no two electrons can have the same set of quantum numbers.

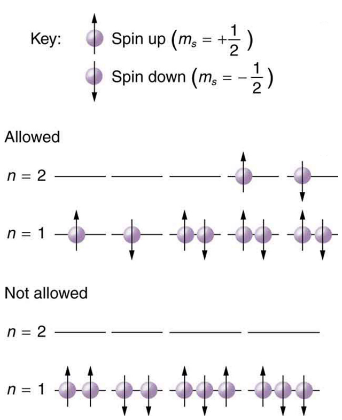
Note:
Pauli Exclusion Principle
No two electrons can have the same set of quantum numbers. That is, no two electrons can be in the same state.

The Austrian physicist Wolfgang Pauli (1900–1958) played a major role in the development of quantum mechanics. He proposed the exclusion principle; hypothesized the existence of an important particle,

called the neutrino, before it was directly observed; made fundamental contributions to several areas of theoretical physics; and influenced many students who went on to do important work of their own. (credit: Nobel Foundation, via Wikimedia Commons)

Let us examine how the exclusion principle applies to electrons in atoms. The quantum numbers involved were defined in Quantum Numbers and Rules as n, l, m_l, s, and m_s. Since s is always $1/2$ for electrons, it is redundant to list s, and so we omit it and specify the state of an electron by a set of four numbers (n, l, m_l, m_s). For example, the quantum numbers $(2, 1, 0, -1/2)$ completely specify the state of an electron in an atom.

Since no two electrons can have the same set of quantum numbers, there are limits to how many of them can be in the same energy state. Note that n determines the energy state in the absence of a magnetic field. So we first choose n, and then we see how many electrons can be in this energy state or energy level. Consider the $n = 1$ level, for example. The only value l can have is 0 (see [link] for a list of possible values once n is known), and thus m_l can only be 0. The spin projection m_s can be either $+1/2$ or $-1/2$, and so there can be two electrons in the $n = 1$ state. One has quantum numbers $(1, 0, 0, +1/2)$, and the other has $(1, 0, 0, -1/2)$. [link] illustrates that there can be one or two electrons having $n = 1$, but not three.



The Pauli exclusion principle explains why some configurations of electrons are allowed while others are not. Since electrons cannot have the same set of quantum numbers, a maximum of two can be

in the $n = 1$ level, and a third electron must reside in the higher-energy $n = 2$ level. If there are two electrons in the $n = 1$ level, their spins must be in opposite directions. (More precisely, their spin projections must differ.)

Shells and Subshells

Because of the Pauli exclusion principle, only hydrogen and helium can have all of their electrons in the $n = 1$ state. Lithium (see the periodic table) has three electrons, and so one must be in the $n = 2$ level. This leads to the concept of shells and shell filling. As we progress up in the number of electrons, we go from hydrogen to helium, lithium, beryllium, boron, and so on, and we see that there are limits to the number of electrons for each value of n. Higher values of the shell n correspond to higher energies, and they can allow more electrons because of the various combinations of l, m_l, and m_s that are possible. Each value of the principal quantum number n thus corresponds to an atomic **shell** into which a limited number of electrons can go. Shells and the number of electrons in them determine the physical and chemical properties of atoms, since it is the outermost electrons that interact most with anything outside the atom.

The probability clouds of electrons with the lowest value of l are closest to the nucleus and, thus, more tightly bound. Thus when shells fill, they start with $l = 0$, progress to $l = 1$, and so on. Each value of l thus corresponds to a **subshell**.

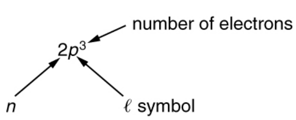
The table given below lists symbols traditionally used to denote shells and subshells.

Shell	**Subshell**	
n	l	***Symbol***
1	0	s
2	1	p
3	2	d
4	3	f

Shell	Subshell	
5	4	g
	5	h
	6[footnote] It is unusual to deal with subshells having l greater than 6, but when encountered, they continue to be labeled in alphabetical order.	i

Shell and Subshell Symbols

To denote shells and subshells, we write $\mathrm{n}l$ with a number for n and a letter for l. For example, an electron in the $n=1$ state must have $l=0$, and it is denoted as a $1s$ electron. Two electrons in the $n=1$ state is denoted as $1s^2$. Another example is an electron in the $n=2$ state with $l=1$, written as $2p$. The case of three electrons with these quantum numbers is written $2p^3$. This notation, called spectroscopic notation, is generalized as shown in [link].



Counting the number of possible combinations of quantum numbers allowed by the exclusion principle, we can determine how many electrons it takes to fill each subshell and shell.

Example:
How Many Electrons Can Be in This Shell?
List all the possible sets of quantum numbers for the $n=2$ shell, and determine the number of electrons that can be in the shell and each of its subshells.

Strategy
Given $n=2$ for the shell, the rules for quantum numbers limit l to be 0 or 1. The shell therefore has two subshells, labeled $2s$ and $2p$. Since the lowest l subshell fills first, we start with the $2s$ subshell possibilities and then proceed with the $2p$ subshell.

Solution
It is convenient to list the possible quantum numbers in a table, as shown below.

n	ℓ	m_l	m_s	Subshell	Total in subshell	Total in shell
2	0	0	+1/2	$2s$	2	8
2	0	0	−1/2			
2	1	1	+1/2	$2p$	6	
2	1	1	−1/2			
2	1	0	+1/2			
2	1	0	−1/2			
2	1	−1	+1/2			
2	1	−1	−1/2			

Discussion
It is laborious to make a table like this every time we want to know how many electrons can be in a shell or subshell. There exist general rules that are easy to apply, as we shall now see.

The number of electrons that can be in a subshell depends entirely on the value of l. Once l is known, there are a fixed number of values of m_l, each of which can have two values for m_s First, since m_l goes from $-l$ to l in steps of 1, there are $2l+1$ possibilities. This number is multiplied by 2, since each electron can be spin up or spin down. Thus the *maximum number of electrons that can be in a subshell* is $2(2l+1)$.

For example, the $2s$ subshell in [link] has a maximum of 2 electrons in it, since $2(2l+1)=2(0+1)=2$ for this subshell. Similarly, the $2p$ subshell has a maximum of 6 electrons, since $2(2l+1)=2(2+1)=6$. For a shell, the maximum number is the sum of what can fit in the subshells. Some algebra shows that the *maximum number of electrons that can be in a shell* is $2n^2$.

For example, for the first shell $n=1$, and so $2n^2=2$. We have already seen that only two electrons can be in the $n=1$ shell. Similarly, for the second shell, $n=2$, and so $2n^2=8$. As found in [link], the total number of electrons in the $n=2$ shell is 8.

Example:
Subshells and Totals for $n=3$
How many subshells are in the $n=3$ shell? Identify each subshell, calculate the maximum number of electrons that will fit into each, and verify that the total is $2n^2$.
Strategy
Subshells are determined by the value of l; thus, we first determine which values of l are allowed, and then we apply the equation “maximum number of electrons that can be in a subshell $=2(2l+1)$” to find the number of electrons in each subshell.
Solution
Since $n=3$, we know that l can be 0, 1, or 2; thus, there are three possible subshells. In standard notation, they are labeled the $3s$, $3p$, and $3d$ subshells. We have already seen that 2 electrons can be in an s state, and 6 in a p state, but let us use the equation “maximum number of electrons that can be in a subshell = $2(2l+1)$” to calculate the maximum number in each:
Equation:

$$\begin{array}{l}3s\text{ has }l=0\text{; thus, }2(2l+1)=2(0+1)=2\\ 3p\text{ has }l=1\text{; thus, }2(2l+1)=2(2+1)=6\\ 3d\text{ has }l=2\text{; thus, }2(2l+1)=2(4+1)=10\\ \text{Total}=18\\ (\text{in the }n=3\text{ shell})\end{array}$$

The equation “maximum number of electrons that can be in a shell = $2n^2$” gives the maximum number in the $n=3$ shell to be
Equation:

$$\text{Maximum number of electrons}=2n^2=2(3)^2=2(9)=18.$$

Discussion
The total number of electrons in the three possible subshells is thus the same as the formula $2n^2$. In standard (spectroscopic) notation, a filled $n=3$ shell is denoted as $3s^2 3p^6 3d^{10}$. Shells do not fill in a simple manner. Before the $n=3$ shell is completely filled, for example, we begin to find electrons in the $n=4$ shell.

Shell Filling and the Periodic Table

[link] shows electron configurations for the first 20 elements in the periodic table, starting with hydrogen and its single electron and ending with calcium. The Pauli exclusion principle determines the maximum number of electrons allowed in each shell and subshell. But the order in which the shells and subshells are filled is complicated because of the large numbers of interactions between electrons.

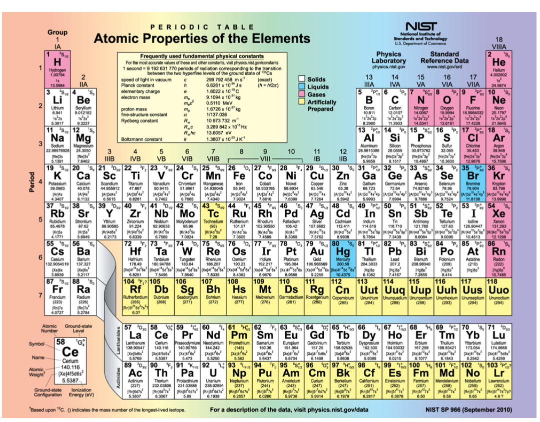
Element	Number of electrons (Z)	Ground state configuration					
H	1	$1s^1$					
He	2	$1s^2$					
Li	3	$1s^2$	$2s^1$				
Be	4	"	$2s^2$				
B	5	"	$2s^2$	$2p^1$			
C	6	"	$2s^2$	$2p^2$			
N	7	"	$2s^2$	$2p^3$			
O	8	"	$2s^2$	$2p^4$			
F	9	"	$2s^2$	$2p^5$			
Ne	10	"	$2s^2$	$2p^6$			
Na	11	"	$2s^2$	$2p^6$	$3s^1$		
Mg	12	"	"	"	$3s^2$		
Al	13	"	"	"	$3s^2$	$3p^1$	
Si	14	"	"	"	$3s^2$	$3p^2$	

Element	Number of electrons (Z)	Ground state configuration					
P	15	"	"	"	$3s^2$	$3p^3$	
S	16	"	"	"	$3s^2$	$3p^4$	
Cl	17	"	"	"	$3s^2$	$3p^5$	
Ar	18	"	"	"	$3s^2$	$3p^6$	
K	19	"	"	"	$3s^2$	$3p^6$	$4s^1$
Ca	20	"	"	"	"	"	$4s^2$

Electron Configurations of Elements Hydrogen Through Calcium

Examining the above table, you can see that as the number of electrons in an atom increases from 1 in hydrogen to 2 in helium and so on, the lowest-energy shell gets filled first—that is, the $n = 1$ shell fills first, and then the $n = 2$ shell begins to fill. Within a shell, the subshells fill starting with the lowest l, or with the s subshell, then the p, and so on, usually until all subshells are filled. The first exception to this occurs for potassium, where the $4s$ subshell begins to fill before any electrons go into the $3d$ subshell. The next exception is not shown in [link]; it occurs for rubidium, where the $5s$ subshell starts to fill before the $4d$ subshell. The reason for these exceptions is that $l = 0$ electrons have probability clouds that penetrate closer to the nucleus and, thus, are more tightly bound (lower in energy).

[link] shows the periodic table of the elements, through element 118. Of special interest are elements in the main groups, namely, those in the columns numbered 1, 2, 13, 14, 15, 16, 17, and 18.



Periodic table of the elements (credit:

National Institute of Standards and Technology, U.S. Department of Commerce)

The number of electrons in the outermost subshell determines the atom’s chemical properties, since it is these electrons that are farthest from the nucleus and thus interact most with other atoms. If the outermost subshell can accept or give up an electron easily, then the atom will be highly reactive chemically. Each group in the periodic table is characterized by its outermost electron configuration. Perhaps the most familiar is Group 18 (Group VIII), the noble gases (helium, neon, argon, etc.). These gases are all characterized by a filled outer subshell that is particularly stable. This means that they have large ionization energies and do not readily give up an electron. Furthermore, if they were to accept an extra electron, it would be in a significantly higher level and thus loosely bound. Chemical reactions often involve sharing electrons. Noble gases can be forced into unstable chemical compounds only under high pressure and temperature.

Group 17 (Group VII) contains the halogens, such as fluorine, chlorine, iodine and bromine, each of which has one less electron than a neighboring noble gas. Each halogen has 5 p electrons (a p^5 configuration), while the p subshell can hold 6 electrons. This means the halogens have one vacancy in their outermost subshell. They thus readily accept an extra electron (it becomes tightly bound, closing the shell as in noble gases) and are highly reactive chemically. The halogens are also likely to form singly negative ions, such as $\mathrm{C1}^-$, fitting an extra electron into the vacancy in the outer subshell. In contrast, alkali metals, such as sodium and potassium, all have a single s electron in their outermost subshell (an s^1 configuration) and are members of Group 1 (Group I). These elements easily give up their extra electron and are thus highly reactive chemically. As you might expect, they also tend to form singly positive ions, such as Na^+, by losing their loosely bound outermost electron. They are metals (conductors), because the loosely bound outer electron can move freely.

Of course, other groups are also of interest. Carbon, silicon, and germanium, for example, have similar chemistries and are in Group 4 (Group IV). Carbon, in particular, is extraordinary in its ability to form many types of bonds and to be part of long chains, such as inorganic molecules. The large group of what are called transitional elements is characterized by the filling of the d subshells and crossing of energy levels. Heavier groups, such as the lanthanide series, are more complex—their shells do not fill in simple order. But the groups recognized by chemists such as Mendeleev have an explanation in the substructure of atoms.

Note:
PhET Explorations: Stern-Gerlach Experiment
Build an atom out of protons, neutrons, and electrons, and see how the element, charge, and mass change. Then play a game to test your ideas!

https://phet.colorado.edu/sims/html/build-an-atom/latest/build-an-atom_en.html

Section Summary

- The state of a system is completely described by a complete set of quantum numbers. This set is written as (n, l, m_l, m_s).
- The Pauli exclusion principle says that no two electrons can have the same set of quantum numbers; that is, no two electrons can be in the same state.
- This exclusion limits the number of electrons in atomic shells and subshells. Each value of n corresponds to a shell, and each value of l corresponds to a subshell.
- The maximum number of electrons that can be in a subshell is $2(2l + 1)$.
- The maximum number of electrons that can be in a shell is $2n^2$.

Conceptual Questions

Exercise:

Problem:

Identify the shell, subshell, and number of electrons for the following: (a) $2p^3$. (b) $4d^9$. (c) $3s^1$. (d) $5g^{16}$.

Exercise:

Problem:

Which of the following are not allowed? State which rule is violated for any that are not allowed. (a) $1p^3$ (b) $2p^8$(c) $3g^{11}$ (d) $4f^2$

Problem Exercises

Exercise:

Problem: (a) How many electrons can be in the $n = 4$ shell?

(b) What are its subshells, and how many electrons can be in each?

Solution:

(a) 32. (b) 2 in s, 6 in p, 10 in d, and 14 in f, for a total of 32.

Exercise:

Problem: (a) What is the minimum value of 1 for a subshell that has 11 electrons in it?

(b) If this subshell is in the $n = 5$ shell, what is the spectroscopic notation for this atom?

Exercise:

Problem:

(a) If one subshell of an atom has 9 electrons in it, what is the minimum value of l? (b) What is the spectroscopic notation for this atom, if this subshell is part of the $n = 3$ shell?

Solution:

(a) 2

(b) $3d^9$

Exercise:

Problem:

(a) List all possible sets of quantum numbers (n, l, m_l, m_s) for the $n = 3$ shell, and determine the number of electrons that can be in the shell and each of its subshells.

(b) Show that the number of electrons in the shell equals $2n^2$ and that the number in each subshell is $2(2l + 1)$.

Exercise:

Problem:

Which of the following spectroscopic notations are not allowed? (a) $5s^1$ (b) $1d^1$ (c) $4s^3$ (d) $3p^7$ (e) $5g^{15}$. State which rule is violated for each that is not allowed.

Solution:

(b) $n \geq l$ is violated,

(c) cannot have 3 electrons in s subshell since $3 > (2l + 1) = 2$

(d) cannot have 7 electrons in p subshell since $7 > (2l + 1) = 2(2 + 1) = 6$

Exercise:

Problem:

Which of the following spectroscopic notations are allowed (that is, which violate none of the rules regarding values of quantum numbers)? (a) $1s^1$(b) $1d^3$(c) $4s^2$ (d) $3p^7$(e) $6h^{20}$

Exercise:

Problem:

(a) Using the Pauli exclusion principle and the rules relating the allowed values of the quantum numbers (n, l, m_l, m_s), prove that the maximum number of electrons in a subshell is $2n^2$.

(b) In a similar manner, prove that the maximum number of electrons in a shell is $2n^2$.

Solution:

(a) The number of different values of m_l is $\pm l, \pm (l - 1), ...,0$ for each $l > 0$ and one for $l = 0 \Rightarrow (2l + 1)$. Also an overall factor of 2 since each m_l can have m_s equal to either $+1/2$ or $-1/2 \Rightarrow 2(2l + 1)$.

(b) for each value of l, you get $2(2l + 1)$

$= 0, 1, 2, ...,(n–1) \Rightarrow 2\{[(2)(0) + 1] + [(2)(1) + 1] + ... + [(2)(n - 1) + 1]\} = 2[\underbrace{1 + 3 + ... + (2n - 3) +}_{n \text{ terms}}$

to see that the expression in the box is $= n^2$, imagine taking $(n - 1)$ from the last term and adding it to first term $= 2[1 + (n–1) + 3 + ... + (2n - 3) + (2n - 1)–(n - 1)] = 2[n + 3 + ... + (2n - 3) + n]$. Now take $(n - 3)$ from penultimate term and add to the second term $2[\underbrace{n + n + ... + n + n}_{n \text{ terms}}] = 2n^2$.

Exercise:

Problem:Integrated Concepts

Estimate the density of a nucleus by calculating the density of a proton, taking it to be a sphere 1.2 fm in diameter. Compare your result with the value estimated in this chapter.

Exercise:

Problem: Integrated Concepts

The electric and magnetic forces on an electron in the CRT in [link] are supposed to be in opposite directions. Verify this by determining the direction of each force for the situation shown. Explain how you obtain the directions (that is, identify the rules used).

Solution:

The electric force on the electron is up (toward the positively charged plate). The magnetic force is down (by the RHR).

Exercise:

Problem:

(a) What is the distance between the slits of a diffraction grating that produces a first-order maximum for the first Balmer line at an angle of $20.0^{\text{º}}$?

(b) At what angle will the fourth line of the Balmer series appear in first order?

(c) At what angle will the second-order maximum be for the first line?

Exercise:

Problem: **Integrated Concepts**

A galaxy moving away from the earth has a speed of $0.0100c$. What wavelength do we observe for an $n_{\text{i}} = 7$ to $n_{\text{f}} = 2$ transition for hydrogen in that galaxy?

Solution:

401 nm

Exercise:

Problem: **Integrated Concepts**

Calculate the velocity of a star moving relative to the earth if you observe a wavelength of 91.0 nm for ionized hydrogen capturing an electron directly into the lowest orbital (that is, a $n_{\text{i}} = \infty$ to $n_{\text{f}} = 1$, or a Lyman series transition).

Exercise:

Problem: **Integrated Concepts**

In a Millikan oil-drop experiment using a setup like that in [link], a 500-V potential difference is applied to plates separated by 2.50 cm. (a) What is the mass of an oil drop having two extra electrons that is suspended motionless by the field between the plates? (b) What is the diameter of the drop, assuming it is a sphere with the density of olive oil?

Solution:

(a) $6.54 \times 10^{-16}\text{ kg}$

(b) $5.54 \times 10^{-7}\text{ m}$

Exercise:

Problem: **Integrated Concepts**

What double-slit separation would produce a first-order maximum at $3.00^{\text{º}}$ for 25.0-keV x rays? The small answer indicates that the wave character of x rays is best determined by having them interact with very small objects such as atoms and molecules.

Exercise:

Problem: **Integrated Concepts**

In a laboratory experiment designed to duplicate Thomson's determination of q_e/m_e, a beam of electrons having a velocity of $6.00 \times 10^{7}\text{ m/s}$ enters a $5.00 \times 10^{-3}\text{ T}$ magnetic field. The beam moves perpendicular

to the field in a path having a 6.80-cm radius of curvature. Determine q_e/m_e from these observations, and compare the result with the known value.

Solution:

$1.76 \times 10^{11}\ \mathrm{C/kg}$, which agrees with the known value of $1.759 \times 10^{11}\ \mathrm{C/kg}$ to within the precision of the measurement

Exercise:

Problem:Integrated Concepts

Find the value of l, the orbital angular momentum quantum number, for the moon around the earth. The extremely large value obtained implies that it is impossible to tell the difference between adjacent quantized orbits for macroscopic objects.

Exercise:

Problem: Integrated Concepts

Particles called muons exist in cosmic rays and can be created in particle accelerators. Muons are very similar to electrons, having the same charge and spin, but they have a mass 207 times greater. When muons are captured by an atom, they orbit just like an electron but with a smaller radius, since the mass in $a_{\mathrm{B}} = \frac{h^2}{4\pi^2 m_e k q_e^2} = 0.529 \times 10^{-10}\ \mathrm{m}$ is 207 m_e.

(a) Calculate the radius of the $n = 1$ orbit for a muon in a uranium ion ($Z = 92$).

(b) Compare this with the 7.5-fm radius of a uranium nucleus. Note that since the muon orbits inside the electron, it falls into a hydrogen-like orbit. Since your answer is less than the radius of the nucleus, you can see that the photons emitted as the muon falls into its lowest orbit can give information about the nucleus.

Solution:

(a) 2.78 fm

(b) 0.37 of the nuclear radius.

Exercise:

Problem:Integrated Concepts

Calculate the minimum amount of energy in joules needed to create a population inversion in a helium-neon laser containing 1.00×10^{-4} moles of neon.

Exercise:

Problem: Integrated Concepts

A carbon dioxide laser used in surgery emits infrared radiation with a wavelength of $10.6\ \mu\mathrm{m}$. In 1.00 ms, this laser raised the temperature of $1.00\ \mathrm{cm}^3$ of flesh to 100ºC and evaporated it.

(a) How many photons were required? You may assume flesh has the same heat of vaporization as water. (b) What was the minimum power output during the flash?

Solution:

(a) 1.34×10^{23}

(b) 2.52 MW

Exercise:

Problem: Integrated Concepts

Suppose an MRI scanner uses 100-MHz radio waves.

(a) Calculate the photon energy.

(b) How does this compare to typical molecular binding energies?

Exercise:

Problem: Integrated Concepts

(a) An excimer laser used for vision correction emits 193-nm UV. Calculate the photon energy in eV.

(b) These photons are used to evaporate corneal tissue, which is very similar to water in its properties. Calculate the amount of energy needed per molecule of water to make the phase change from liquid to gas. That is, divide the heat of vaporization in kJ/kg by the number of water molecules in a kilogram.

(c) Convert this to eV and compare to the photon energy. Discuss the implications.

Solution:

(a) 6.42 eV

(b) $7.27 \times 10^{-20}\text{ J/molecule}$

(c) 0.454 eV, 14.1 times less than a single UV photon. Therefore, each photon will evaporate approximately 14 molecules of tissue. This gives the surgeon a rather precise method of removing corneal tissue from the surface of the eye.

Exercise:

Problem: Integrated Concepts

A neighboring galaxy rotates on its axis so that stars on one side move toward us as fast as 200 km/s, while those on the other side move away as fast as 200 km/s. This causes the EM radiation we receive to be Doppler shifted by velocities over the entire range of ±200 km/s. What range of wavelengths will we observe for the 656.0-nm line in the Balmer series of hydrogen emitted by stars in this galaxy. (This is called line broadening.)

Exercise:

Problem: Integrated Concepts

A pulsar is a rapidly spinning remnant of a supernova. It rotates on its axis, sweeping hydrogen along with it so that hydrogen on one side moves toward us as fast as 50.0 km/s, while that on the other side moves away as fast as 50.0 km/s. This means that the EM radiation we receive will be Doppler shifted over a range of $\pm 50.0\text{ km/s}$. What range of wavelengths will we observe for the 91.20-nm line in the Lyman series of hydrogen? (Such line broadening is observed and actually provides part of the evidence for rapid rotation.)

Solution:

91.18 nm to 91.22 nm

Exercise:

Problem: **Integrated Concepts**

Prove that the velocity of charged particles moving along a straight path through perpendicular electric and magnetic fields is $v = E/B$. Thus crossed electric and magnetic fields can be used as a velocity selector independent of the charge and mass of the particle involved.

Exercise:

Problem: **Unreasonable Results**

(a) What voltage must be applied to an X-ray tube to obtain 0.0100-fm-wavelength X-rays for use in exploring the details of nuclei? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) 1.24×10^{11} V

(b) The voltage is extremely large compared with any practical value.

(c) The assumption of such a short wavelength by this method is unreasonable.

Exercise:

Problem: **Unreasonable Results**

A student in a physics laboratory observes a hydrogen spectrum with a diffraction grating for the purpose of measuring the wavelengths of the emitted radiation. In the spectrum, she observes a yellow line and finds its wavelength to be 589 nm. (a) Assuming this is part of the Balmer series, determine n_{i}, the principal quantum number of the initial state. (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Exercise:

Problem: **Construct Your Own Problem**

The solar corona is so hot that most atoms in it are ionized. Consider a hydrogen-like atom in the corona that has only a single electron. Construct a problem in which you calculate selected spectral energies and wavelengths of the Lyman, Balmer, or other series of this atom that could be used to identify its presence in a very hot gas. You will need to choose the atomic number of the atom, identify the element, and choose which spectral lines to consider.

Exercise:

Problem: **Construct Your Own Problem**

Consider the Doppler-shifted hydrogen spectrum received from a rapidly receding galaxy. Construct a problem in which you calculate the energies of selected spectral lines in the Balmer series and examine whether they can be described with a formula like that in the equation $\frac{1}{\lambda} = R\left(\frac{1}{n_{\mathrm{f}}^2} - \frac{1}{n_{\mathrm{i}}^2}\right)$, but with a different constant R.

Glossary

atomic number
: the number of protons in the nucleus of an atom

Pauli exclusion principle

a principle that states that no two electrons can have the same set of quantum numbers; that is, no two electrons can be in the same state

shell

a probability cloud for electrons that has a single principal quantum number

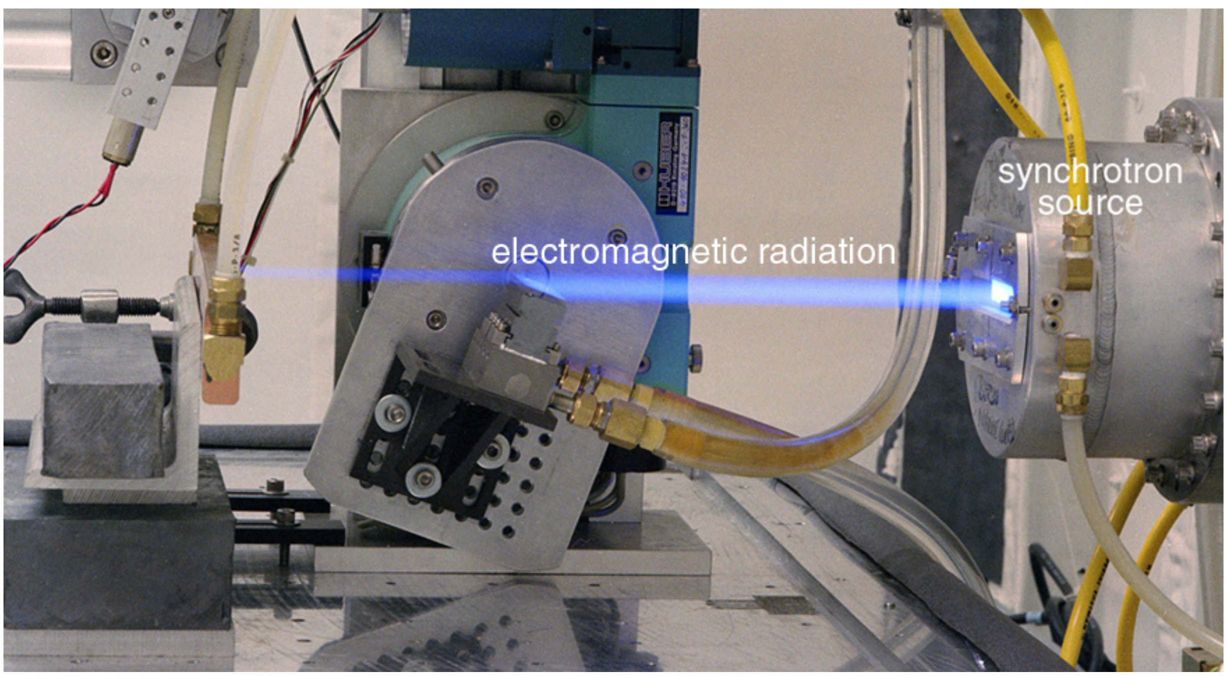
subshell

the probability cloud for electrons that has a single angular momentum quantum number l

Introduction to Radioactivity and Nuclear Physics class="introduction"

- Define radioactivity.

The synchrotron source produces electromagnetic radiation, as evident from the visible glow. (credit: United States Department of Energy, via Wikimedia Commons)

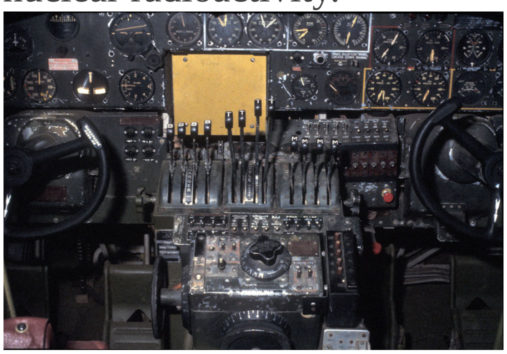


There is an ongoing quest to find substructures of matter. At one time, it was thought that atoms would be the ultimate substructure, but just when the first direct evidence of atoms was obtained, it became clear that they have a substructure and a tiny *nucleus*. The nucleus itself has spectacular characteristics. For example, certain nuclei are unstable, and their decay emits radiations with energies millions of times greater than atomic energies. Some of the mysteries of nature, such as why the core of the earth remains molten and how the sun produces its energy, are explained by nuclear phenomena. The exploration of *radioactivity* and the nucleus revealed fundamental and previously unknown particles, forces, and conservation laws. That exploration has evolved into a search for further underlying structures, such as quarks. In this chapter, the fundamentals of nuclear radioactivity and the nucleus are explored. The following two chapters explore the more important applications of nuclear physics in the field of medicine. We will also explore the basics of what we know about quarks and other substructures smaller than nuclei.

Nuclear Radioactivity

- Explain nuclear radiation.
- Explain the types of radiation—alpha emission, beta emission, and gamma emission.
- Explain the ionization of radiation in an atom.
- Define the range of radiation.

The discovery and study of nuclear radioactivity quickly revealed evidence of revolutionary new physics. In addition, uses for nuclear radiation also emerged quickly—for example, people such as Ernest Rutherford used it to determine the size of the nucleus and devices were painted with radon-doped paint to make them glow in the dark (see [link]). We therefore begin our study of nuclear physics with the discovery and basic features of nuclear radioactivity.



The dials of this World War II aircraft glow in the dark, because they are painted with radium-doped phosphorescent paint. It is a poignant reminder of the dual nature of radiation. Although radium paint dials are conveniently visible day and night, they emit radon, a radioactive gas that is hazardous and is not

directly sensed. (credit: U.S. Air Force Photo)

Discovery of Nuclear Radioactivity

In 1896, the French physicist Antoine Henri Becquerel (1852–1908) accidentally found that a uranium-rich mineral called pitchblende emits invisible, penetrating rays that can darken a photographic plate enclosed in an opaque envelope. The rays therefore carry energy; but amazingly, the pitchblende emits them continuously without any energy input. This is an apparent violation of the law of conservation of energy, one that we now understand is due to the conversion of a small amount of mass into energy, as related in Einstein's famous equation $E = mc^2$. It was soon evident that Becquerel's rays originate in the nuclei of the atoms and have other unique characteristics. The emission of these rays is called **nuclear radioactivity** or simply **radioactivity**. The rays themselves are called **nuclear radiation**. A nucleus that spontaneously destroys part of its mass to emit radiation is said to **decay** (a term also used to describe the emission of radiation by atoms in excited states). A substance or object that emits nuclear radiation is said to be **radioactive**.

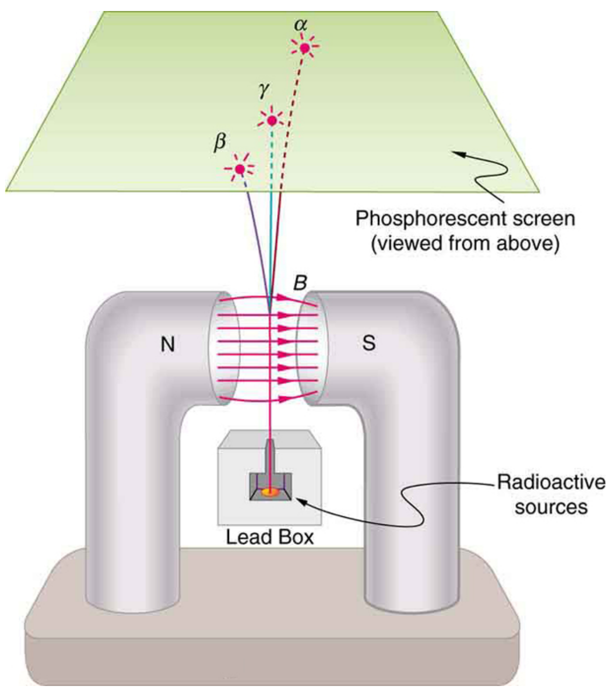
Two types of experimental evidence imply that Becquerel's rays originate deep in the heart (or nucleus) of an atom. First, the radiation is found to be associated with certain elements, such as uranium. Radiation does not vary with chemical state—that is, uranium is radioactive whether it is in the form of an element or compound. In addition, radiation does not vary with temperature, pressure, or ionization state of the uranium atom. Since all of these factors affect electrons in an atom, the radiation cannot come from electron transitions, as atomic spectra do. The huge energy emitted during each event is the second piece of evidence that the radiation cannot be atomic. Nuclear radiation has energies of the order of 10^6 eV per event, which is much greater than the typical atomic energies (a few eV), such as that observed in spectra and chemical reactions, and more than ten times as high as the most energetic characteristic x rays. Becquerel did not vigorously pursue his discovery for very long. In 1898, Marie Curie (1867–

1934), then a graduate student married the already well-known French physicist Pierre Curie (1859–1906), began her doctoral study of Becquerel's rays. She and her husband soon discovered two new radioactive elements, which she named *polonium* (after her native land) and *radium* (because it radiates). These two new elements filled holes in the periodic table and, further, displayed much higher levels of radioactivity per gram of material than uranium. Over a period of four years, working under poor conditions and spending their own funds, the Curies processed more than a ton of uranium ore to isolate a gram of radium salt. Radium became highly sought after, because it was about two million times as radioactive as uranium. Curie's radium salt glowed visibly from the radiation that took its toll on them and other unaware researchers. Shortly after completing her Ph.D., both Curies and Becquerel shared the 1903 Nobel Prize in physics for their work on radioactivity. Pierre was killed in a horse cart accident in 1906, but Marie continued her study of radioactivity for nearly 30 more years. Awarded the 1911 Nobel Prize in chemistry for her discovery of two new elements, she remains the only person to win Nobel Prizes in physics and chemistry. Marie's radioactive fingerprints on some pages of her notebooks can still expose film, and she suffered from radiation-induced lesions. She died of leukemia likely caused by radiation, but she was active in research almost until her death in 1934. The following year, her daughter and son-in-law, Irene and Frederic Joliot-Curie, were awarded the Nobel Prize in chemistry for their discovery of artificially induced radiation, adding to a remarkable family legacy.

Alpha, Beta, and Gamma

Research begun by people such as New Zealander Ernest Rutherford soon after the discovery of nuclear radiation indicated that different types of rays are emitted. Eventually, three types were distinguished and named **alpha** (α), **beta**(β), and **gamma**(γ), because, like x-rays, their identities were initially unknown. [link] shows what happens if the rays are passed through a magnetic field. The γs are unaffected, while the α s and β s are deflected in opposite directions, indicating the α s are positive, the β s negative, and the γ s uncharged. Rutherford used both magnetic and electric fields to show that α s have a positive charge twice the magnitude of an electron, or $+2 \mid q_e \mid$. In the process, he found the α s charge to mass ratio to be several

thousand times smaller than the electron's. Later on, Rutherford collected α s from a radioactive source and passed an electric discharge through them, obtaining the spectrum of recently discovered helium gas. Among many important discoveries made by Rutherford and his collaborators was the proof that α *radiation is the emission of a helium nucleus*. Rutherford won the Nobel Prize in chemistry in 1908 for his early work. He continued to make important contributions until his death in 1934.



Alpha, beta, and gamma rays are passed through a magnetic field on the way to a phosphorescent screen. The α s and β s bend in opposite directions, while the γ s are unaffected, indicating a positive charge for α s, negative for β s, and neutral for γ s. Consistent results are obtained with electric fields. Collection of the radiation offers further

confirmation from the direct measurement of excess charge.

Other researchers had already proved that β s are negative and have the same mass and same charge-to-mass ratio as the recently discovered electron. By 1902, it was recognized that β *radiation is the emission of an electron*. Although β s are electrons, they do not exist in the nucleus before it decays and are not ejected atomic electrons—the electron is created in the nucleus at the instant of decay.

Since γ s remain unaffected by electric and magnetic fields, it is natural to think they might be photons. Evidence for this grew, but it was not until 1914 that this was proved by Rutherford and collaborators. By scattering γ radiation from a crystal and observing interference, they demonstrated that γ *radiation is the emission of a high-energy photon by a nucleus*. In fact, γ radiation comes from the de-excitation of a nucleus, just as an x ray comes from the de-excitation of an atom. The names "γ ray" and "x ray" identify the source of the radiation. At the same energy, γ rays and x rays are otherwise identical.

Type of Radiation	**Range**
α -Particles	A sheet of paper, a few cm of air, fractions of a mm of tissue

Type of Radiation	Range
β -Particles	A thin aluminum plate, or tens of cm of tissue
γ Rays	Several cm of lead or meters of concrete

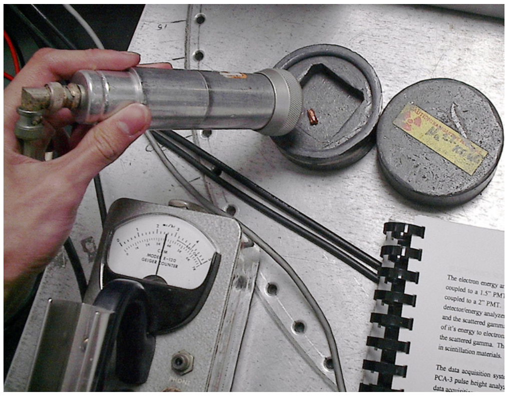
Properties of Nuclear Radiation

Ionization and Range

Two of the most important characteristics of α, β, and γ rays were recognized very early. All three types of nuclear radiation produce *ionization* in materials, but they penetrate different distances in materials—that is, they have different *ranges*. Let us examine why they have these characteristics and what are some of the consequences.

Like x rays, nuclear radiation in the form of α s, β s, and γ s has enough energy per event to ionize atoms and molecules in any material. The energy emitted in various nuclear decays ranges from a few keV to more than 10 MeV, while only a few eV are needed to produce ionization. The effects of x rays and nuclear radiation on biological tissues and other materials, such as solid state electronics, are directly related to the ionization they produce. All of them, for example, can damage electronics or kill cancer cells. In addition, methods for detecting x rays and nuclear radiation are based on ionization, directly or indirectly. All of them can ionize the air between the plates of a capacitor, for example, causing it to discharge. This is the basis of inexpensive personal radiation monitors, such as pictured in [link]. Apart from α, β, and γ, there are other forms of nuclear radiation as well, and these also produce ionization with similar effects. We define **ionizing radiation** as any form of radiation that produces ionization

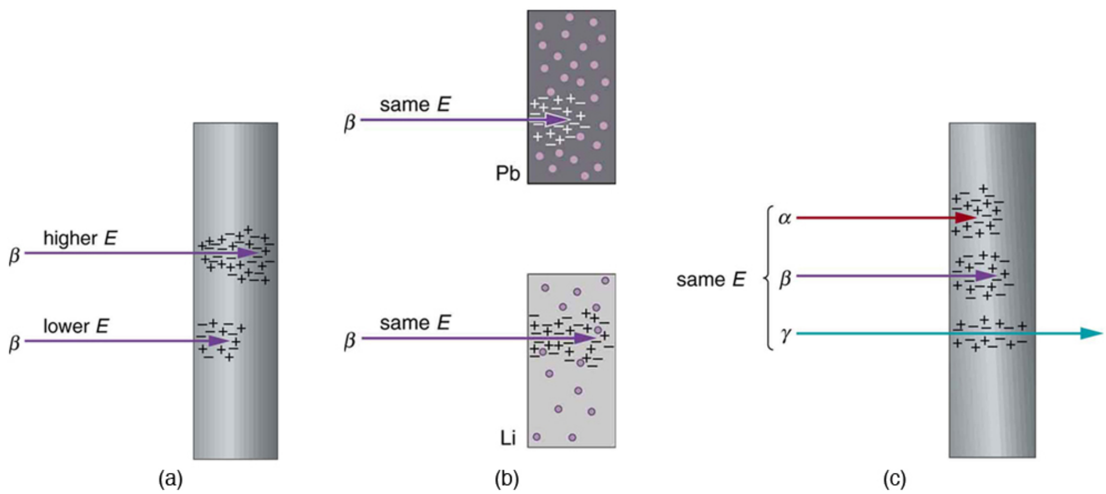
whether nuclear in origin or not, since the effects and detection of the radiation are related to ionization.



These dosimeters (literally, dose meters) are personal radiation monitors that detect the amount of radiation by the discharge of a rechargeable internal capacitor. The amount of discharge is related to the amount of ionizing radiation encountered, a measurement of dose. One dosimeter is shown in the charger. Its scale is read through an eyepiece on the top. (credit: L. Chang, Wikimedia Commons)

The **range of radiation** is defined to be the distance it can travel through a material. Range is related to several factors, including the energy of the

radiation, the material encountered, and the type of radiation (see [link]). The higher the *energy*, the greater the range, all other factors being the same. This makes good sense, since radiation loses its energy in materials primarily by producing ionization in them, and each ionization of an atom or a molecule requires energy that is removed from the radiation. The amount of ionization is, thus, directly proportional to the energy of the particle of radiation, as is its range.



The penetration or range of radiation depends on its energy, the material it encounters, and the type of radiation. (a) Greater energy means greater range. (b) Radiation has a smaller range in materials with high electron density. (c) Alphas have the smallest range, betas have a greater range, and gammas penetrate the farthest.

Radiation can be absorbed or shielded by materials, such as the lead aprons dentists drape on us when taking x rays. Lead is a particularly effective shield compared with other materials, such as plastic or air. How does the range of radiation depend on *material*? Ionizing radiation interacts best with charged particles in a material. Since electrons have small masses, they most readily absorb the energy of the radiation in collisions. The greater the

density of a material and, in particular, the greater the density of electrons within a material, the smaller the range of radiation.

Note:
Collisions
Conservation of energy and momentum often results in energy transfer to a less massive object in a collision. This was discussed in detail in Work, Energy, and Energy Resources, for example.

Different *types* of radiation have different ranges when compared at the same energy and in the same material. Alphas have the shortest range, betas penetrate farther, and gammas have the greatest range. This is directly related to charge and speed of the particle or type of radiation. At a given energy, each α, β, or γ will produce the same number of ionizations in a material (each ionization requires a certain amount of energy on average). The more readily the particle produces ionization, the more quickly it will lose its energy. The effect of *charge* is as follows: The α has a charge of $+2q_e$, the β has a charge of $-q_e$, and the γ is uncharged. The electromagnetic force exerted by the α is thus twice as strong as that exerted by the β and it is more likely to produce ionization. Although chargeless, the γ does interact weakly because it is an electromagnetic wave, but it is less likely to produce ionization in any encounter. More quantitatively, the change in momentum Δp given to a particle in the material is $\Delta p = F\Delta t$, where F is the force the α, β, or γ exerts over a time Δt. The smaller the charge, the smaller is F and the smaller is the momentum (and energy) lost. Since the speed of alphas is about 5% to 10% of the speed of light, classical (non-relativistic) formulas apply.

The *speed* at which they travel is the other major factor affecting the range of α s, β s, and γ s. The faster they move, the less time they spend in the vicinity of an atom or a molecule, and the less likely they are to interact. Since α s and β s are particles with mass (helium nuclei and electrons, respectively), their energy is kinetic, given classically by $\frac{1}{2}mv^2$. The mass

of the β particle is thousands of times less than that of the α s, so that β s must travel much faster than α s to have the same energy. Since β s move faster (most at relativistic speeds), they have less time to interact than α s. Gamma rays are photons, which must travel at the speed of light. They are even less likely to interact than a β, since they spend even less time near a given atom (and they have no charge). The range of γ s is thus greater than the range of β s.

Alpha radiation from radioactive sources has a range much less than a millimeter of biological tissues, usually not enough to even penetrate the dead layers of our skin. On the other hand, the same α radiation can penetrate a few centimeters of air, so mere distance from a source prevents α radiation from reaching us. This makes α radiation relatively safe for our body compared to β and γ radiation. Typical β radiation can penetrate a few millimeters of tissue or about a meter of air. Beta radiation is thus hazardous even when not ingested. The range of β s in lead is about a millimeter, and so it is easy to store β sources in lead radiation-proof containers. Gamma rays have a much greater range than either αs or βs. In fact, if a given thickness of material, like a lead brick, absorbs 90% of the γ s, then a second lead brick will only absorb 90% of what got through the first. Thus, γs do not have a well-defined range; we can only cut down the amount that gets through. Typically, γs can penetrate many meters of air, go right through our bodies, and are effectively shielded (that is, reduced in intensity to acceptable levels) by many centimeters of lead. One benefit of γ s is that they can be used as radioactive tracers (see [link]).

This image of the concentration of a radioactive tracer in a patient's body reveals where the most active bone cells are, an indication of bone cancer. A short-lived radioactive substance that locates itself selectively is given to the patient, and the radiation is measured with an external detector. The emitted γ radiation has a sufficient range to leave the body—the range of α s and β s is too small for them to be observed outside the patient. (credit: Kieran Maher, Wikimedia Commons)

Note:
PhET Explorations: Beta Decay
Build an atom out of protons, neutrons, and electrons, and see how the element, charge, and mass change. Then play a game to test your ideas!

https://archive.cnx.org/specials/f0a27b96-f5c8-11e5-a22c-73f8c149bebf/beta-decay/#sim-multiple-atoms

Section Summary

- Some nuclei are radioactive—they spontaneously decay destroying some part of their mass and emitting energetic rays, a process called nuclear radioactivity.
- Nuclear radiation, like x rays, is ionizing radiation, because energy sufficient to ionize matter is emitted in each decay.
- The range (or distance traveled in a material) of ionizing radiation is directly related to the charge of the emitted particle and its energy, with greater-charge and lower-energy particles having the shortest ranges.
- Radiation detectors are based directly or indirectly upon the ionization created by radiation, as are the effects of radiation on living and inert materials.

Conceptual Questions

Exercise:

Problem:

Suppose the range for $5.0\ \text{MeV}\alpha$ ray is known to be 2.0 mm in a certain material. Does this mean that every $5.0\ \text{MeV}\alpha$ a ray that strikes this material travels 2.0 mm, or does the range have an average value with some statistical fluctuations in the distances traveled? Explain.

Exercise:

Problem:

What is the difference between γ rays and characteristic x rays? Is either necessarily more energetic than the other? Which can be the most energetic?

Exercise:

Problem:

Ionizing radiation interacts with matter by scattering from electrons and nuclei in the substance. Based on the law of conservation of momentum and energy, explain why electrons tend to absorb more energy than nuclei in these interactions.

Exercise:

Problem:

What characteristics of radioactivity show it to be nuclear in origin and not atomic?

Exercise:

Problem:

What is the source of the energy emitted in radioactive decay? Identify an earlier conservation law, and describe how it was modified to take such processes into account.

Exercise:

Problem:

Consider [link]. If an electric field is substituted for the magnetic field with positive charge instead of the north pole and negative charge instead of the south pole, in which directions will the α, β, and γ rays bend?

Exercise:

Problem:

Explain how an α particle can have a larger range in air than a β particle with the same energy in lead.

Exercise:

Problem:

Arrange the following according to their ability to act as radiation shields, with the best first and worst last. Explain your ordering in terms of how radiation loses its energy in matter.

(a) A solid material with low density composed of low-mass atoms.

(b) A gas composed of high-mass atoms.

(c) A gas composed of low-mass atoms.

(d) A solid with high density composed of high-mass atoms.

Exercise:

Problem:

Often, when people have to work around radioactive materials spills, we see them wearing white coveralls (usually a plastic material). What types of radiation (if any) do you think these suits protect the worker from, and how?

Glossary

alpha rays
: one of the types of rays emitted from the nucleus of an atom

beta rays
: one of the types of rays emitted from the nucleus of an atom

gamma rays

one of the types of rays emitted from the nucleus of an atom

ionizing radiation
radiation (whether nuclear in origin or not) that produces ionization whether nuclear in origin or not

nuclear radiation
rays that originate in the nuclei of atoms, the first examples of which were discovered by Becquerel

radioactivity
the emission of rays from the nuclei of atoms

radioactive
a substance or object that emits nuclear radiation

range of radiation
the distance that the radiation can travel through a material

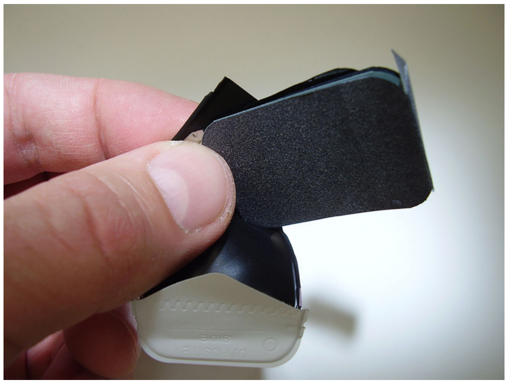
Radiation Detection and Detectors

- Explain the working principle of a Geiger tube.
- Define and discuss radiation detectors.

It is well known that ionizing radiation affects us but does not trigger nerve impulses. Newspapers carry stories about unsuspecting victims of radiation poisoning who fall ill with radiation sickness, such as burns and blood count changes, but who never felt the radiation directly. This makes the detection of radiation by instruments more than an important research tool. This section is a brief overview of radiation detection and some of its applications.

Human Application

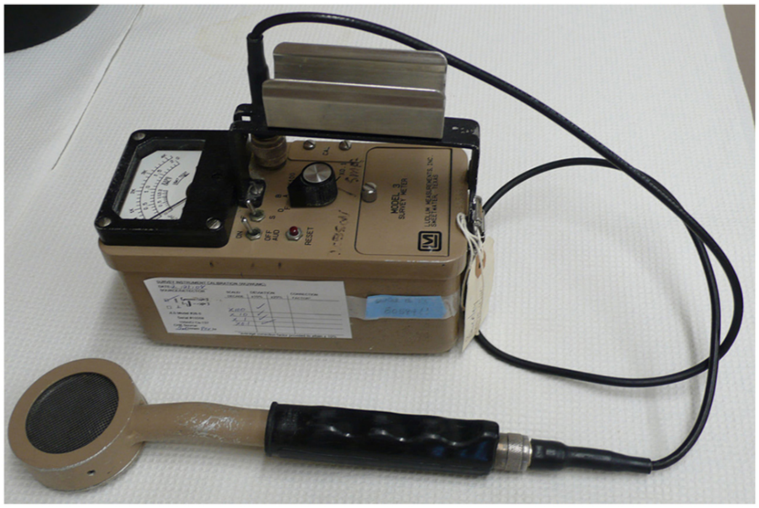
The first direct detection of radiation was Becquerel's fogged photographic plate. Photographic film is still the most common detector of ionizing radiation, being used routinely in medical and dental x rays. Nuclear radiation is also captured on film, such as seen in [link]. The mechanism for film exposure by ionizing radiation is similar to that by photons. A quantum of energy interacts with the emulsion and alters it chemically, thus exposing the film. The quantum come from an α-particle, β-particle, or photon, provided it has more than the few eV of energy needed to induce the chemical change (as does all ionizing radiation). The process is not 100% efficient, since not all incident radiation interacts and not all interactions produce the chemical change. The amount of film darkening is related to exposure, but the darkening also depends on the type of radiation, so that absorbers and other devices must be used to obtain energy, charge, and particle-identification information.



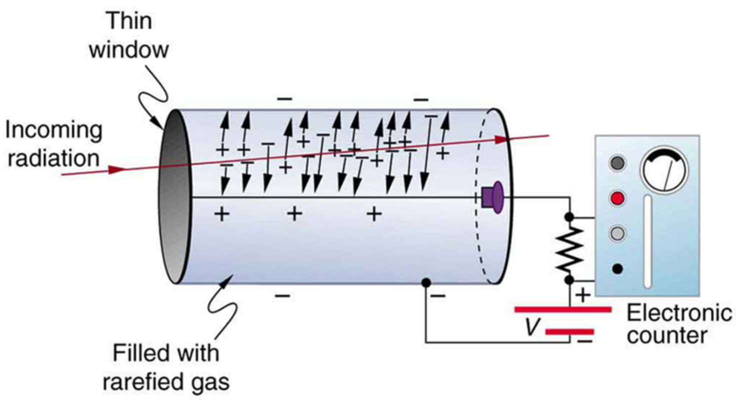
Film badges contain film similar to that used in this dental x-ray film and is sandwiched between various absorbers to determine the penetrating ability of the radiation as well as the amount. (credit: Werneuchen, Wikimedia Commons)

Another very common **radiation detector** is the **Geiger tube**. The clicking and buzzing sound we hear in dramatizations and documentaries, as well as in our own physics labs, is usually an audio output of events detected by a Geiger counter. These relatively inexpensive radiation detectors are based on the simple and sturdy Geiger tube, shown schematically in [link](b). A conducting cylinder with a wire along its axis is filled with an insulating gas so that a voltage applied between the cylinder and wire produces almost no current. Ionizing radiation passing through the tube produces free ion pairs that are attracted to the wire and cylinder, forming a current that is detected as a count. The word count implies that there is no information on energy, charge, or type of radiation with a simple Geiger counter. They do not detect every particle, since some radiation can pass through without producing enough ionization to be detected. However, Geiger counters are very useful in producing a prompt output that reveals the existence and relative intensity of ionizing radiation.

(a)



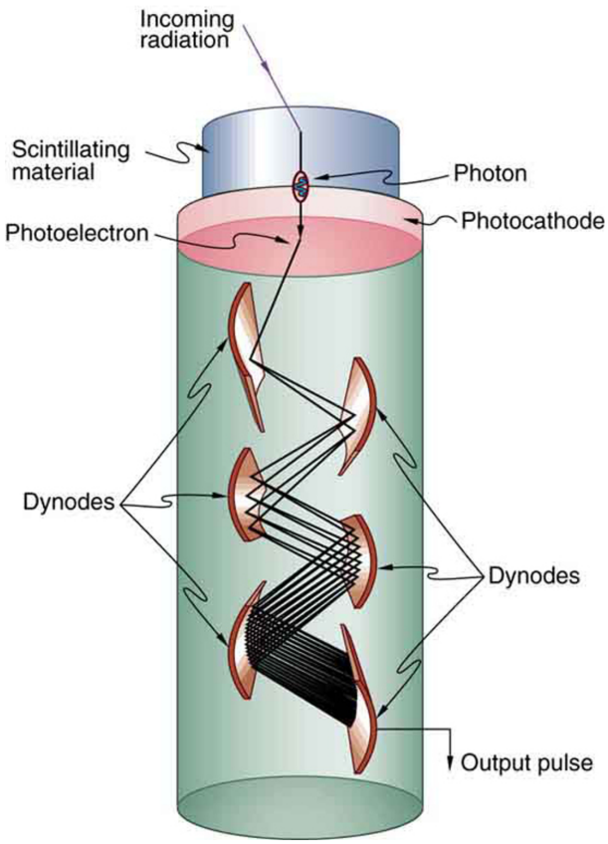
(b)



(a) Geiger counters such as this one are used for prompt monitoring of radiation levels, generally giving only relative intensity and not identifying the type or energy of the radiation. (credit: TimVickers, Wikimedia Commons) (b) Voltage applied between the cylinder and wire in a Geiger tube causes ions and electrons produced by radiation passing through the gas-filled cylinder to move towards them. The resulting current is detected and registered as a count.

Another radiation detection method records light produced when radiation interacts with materials. The energy of the radiation is sufficient to excite atoms in a material that may fluoresce, such as the phosphor used by Rutherford’s group. Materials called **scintillators** use a more complex collaborative process to convert radiation energy into light. Scintillators may be liquid or solid, and they can be very efficient. Their light output can provide information about the energy, charge, and type of radiation. Scintillator light flashes are very brief in duration, enabling the detection of a huge number of particles in short periods of time. Scintillator detectors are used in a variety of research and diagnostic applications. Among these are the detection by satellite-mounted equipment of the radiation from distant galaxies, the analysis of radiation from a person indicating body burdens, and the detection of exotic particles in accelerator laboratories.

Light from a scintillator is converted into electrical signals by devices such as the **photomultiplier** tube shown schematically in [link]. These tubes are based on the photoelectric effect, which is multiplied in stages into a cascade of electrons, hence the name photomultiplier. Light entering the photomultiplier strikes a metal plate, ejecting an electron that is attracted by a positive potential difference to the next plate, giving it enough energy to eject two or more electrons, and so on. The final output current can be made proportional to the energy of the light entering the tube, which is in turn proportional to the energy deposited in the scintillator. Very sophisticated information can be obtained with scintillators, including energy, charge, particle identification, direction of motion, and so on.



Photomultipliers use the photoelectric effect on the photocathode to convert the light output of a scintillator into an electrical signal. Each successive dynode has a more-positive potential than the last and attracts the ejected electrons, giving them more energy. The number of electrons is thus multiplied at each dynode, resulting in an easily detected output current.

Solid-state radiation detectors convert ionization produced in a semiconductor (like those found in computer chips) directly into an

electrical signal. Semiconductors can be constructed that do not conduct current in one particular direction. When a voltage is applied in that direction, current flows only when ionization is produced by radiation, similar to what happens in a Geiger tube. Further, the amount of current in a solid-state detector is closely related to the energy deposited and, since the detector is solid, it can have a high efficiency (since ionizing radiation is stopped in a shorter distance in solids fewer particles escape detection). As with scintillators, very sophisticated information can be obtained from solid-state detectors.

Note:
PhET Explorations: Radioactive Dating Game
Learn about different types of radiometric dating, such as carbon dating. Understand how decay and half life work to enable radiometric dating to work. Play a game that tests your ability to match the percentage of the dating element that remains to the age of the object.

https://archive.cnx.org/specials/d709a8b0-068c-11e6-bcfb-f38266817c66/radioactive-dating-game/#sim-half-life

Section Summary

- Radiation detectors are based directly or indirectly upon the ionization created by radiation, as are the effects of radiation on living and inert materials.

Conceptual Questions

Exercise:

Problem:

Is it possible for light emitted by a scintillator to be too low in frequency to be used in a photomultiplier tube? Explain.

Problems & Exercises

Exercise:

Problem:

The energy of 30.0 eV is required to ionize a molecule of the gas inside a Geiger tube, thereby producing an ion pair. Suppose a particle of ionizing radiation deposits 0.500 MeV of energy in this Geiger tube. What maximum number of ion pairs can it create?

Solution:

1.67×10^4

Exercise:

Problem:

A particle of ionizing radiation creates 4000 ion pairs in the gas inside a Geiger tube as it passes through. What minimum energy was deposited, if 30.0 eV is required to create each ion pair?

Exercise:

Problem:

(a) Repeat [link], and convert the energy to joules or calories. (b) If all of this energy is converted to thermal energy in the gas, what is its temperature increase, assuming $50.0\ \text{cm}^3$ of ideal gas at 0.250-atm pressure? (The small answer is consistent with the fact that the energy is large on a quantum mechanical scale but small on a macroscopic scale.)

Exercise:

Problem:

Suppose a particle of ionizing radiation deposits 1.0 MeV in the gas of a Geiger tube, all of which goes to creating ion pairs. Each ion pair requires 30.0 eV of energy. (a) The applied voltage sweeps the ions out of the gas in 1.00 μs. What is the current? (b) This current is smaller than the actual current since the applied voltage in the Geiger tube accelerates the separated ions, which then create other ion pairs in subsequent collisions. What is the current if this last effect multiplies the number of ion pairs by 900?

Glossary

Geiger tube
: a very common radiation detector that usually gives an audio output

photomultiplier
: a device that converts light into electrical signals

radiation detector
: a device that is used to detect and track the radiation from a radioactive reaction

scintillators
: a radiation detection method that records light produced when radiation interacts with materials

solid-state radiation detectors
: semiconductors fabricated to directly convert incident radiation into electrical current

Substructure of the Nucleus

- Define and discuss the nucleus in an atom.
- Define atomic number.
- Define and discuss isotopes.
- Calculate the density of the nucleus.
- Explain nuclear force.

What is inside the nucleus? Why are some nuclei stable while others decay? (See [link].) Why are there different types of decay (α, β and γ)? Why are nuclear decay energies so large? Pursuing natural questions like these has led to far more fundamental discoveries than you might imagine.

(a) (b) (c)

Why is most of the carbon in this coal stable (a), while the uranium in the disk (b) slowly decays over billions of years? Why is cesium in this ampule (c) even less stable than the uranium, decaying in far less than 1/1,000,000 the time? What is the reason uranium and cesium undergo different types of decay (α and β, respectively)? (credits: (a) Bresson Thomas, Wikimedia Commons; (b) U.S. Department of Energy; (c) Tomihahndorf, Wikimedia Commons)

We have already identified **protons** as the particles that carry positive charge in the nuclei. However, there are actually *two* types of particles in the nuclei—the *proton* and the *neutron*, referred to collectively as **nucleons**, the constituents of nuclei. As its name implies, the **neutron** is a neutral particle ($q = 0$) that has

nearly the same mass and intrinsic spin as the proton. [link] compares the masses of protons, neutrons, and electrons. Note how close the proton and neutron masses are, but the neutron is slightly more massive once you look past the third digit. Both nucleons are much more massive than an electron. In fact, $m_p = 1836 m_e$ (as noted in Medical Applications of Nuclear Physics and $m_n = 1839 m_e$.

[link] also gives masses in terms of mass units that are more convenient than kilograms on the atomic and nuclear scale. The first of these is the *unified* ***atomic mass*** *unit* (u), defined as

Equation:

$$1 \text{ u} = 1.6605 \times 10^{-27} \text{ kg}.$$

This unit is defined so that a neutral carbon ^{12}C atom has a mass of exactly 12 u. Masses are also expressed in units of MeV/c^2. These units are very convenient when considering the conversion of mass into energy (and vice versa), as is so prominent in nuclear processes. Using $E = mc^2$ and units of m in MeV/c^2, we find that c^2 cancels and E comes out conveniently in MeV. For example, if the rest mass of a proton is converted entirely into energy, then

Equation:

$$E = mc^2 = (938.27 \text{ MeV}/c^2)c^2 = 938.27 \text{ MeV}.$$

It is useful to note that 1 u of mass converted to energy produces 931.5 MeV, or

Equation:

$$1 \text{ u} = 931.5 \text{ MeV}/c^2.$$

All properties of a nucleus are determined by the number of protons and neutrons it has. A specific combination of protons and neutrons is called a **nuclide** and is a unique nucleus. The following notation is used to represent a particular nuclide:

Equation:

$$^{A}_{Z}\text{X}_N,$$

where the symbols A, X, Z , and N are defined as follows: The *number of protons in a nucleus* is the **atomic number** Z, as defined in Medical Applications of Nuclear Physics. X is the *symbol for the element*, such as Ca for calcium. However, once Z is known, the element is known; hence, Z and X are redundant. For example, $Z = 20$ is always calcium, and calcium always has $Z = 20$. N is the *number of neutrons* in a nucleus. In the notation for a nuclide, the subscript N is usually omitted. The symbol A is defined as the number of nucleons or the *total number of protons and neutrons*,

Equation:

$$A = N + Z,$$

where A is also called the **mass number**. This name for A is logical; the mass of an atom is nearly equal to the mass of its nucleus, since electrons have so little mass. The mass of the nucleus turns out to be nearly equal to the sum of the masses of the protons and neutrons in it, which is proportional to A. In this context, it is particularly convenient to express masses in units of u. Both protons and neutrons have masses close to 1 u, and so the mass of an atom is close to A u. For example, in an oxygen nucleus with eight protons and eight neutrons, $A = 16$, and its mass is 16 u. As noticed, the unified atomic mass unit is defined so that a neutral carbon atom (actually a ^{12}C atom) has a mass of *exactly* 12 u. Carbon was chosen as the standard, partly because of its importance in organic chemistry (see Appendix A).

Particle	**Symbol**	**kg**	**u**	**MeV*c*^{2}**
Proton	*p*	1.67262×10^{-27}	1.007276	938.27
Neutron	*n*	1.67493×10^{-27}	1.008665	939.57

Particle	Symbol	kg	u	MeVc^2
Electron	*e*	9.1094×10^{-31}	0.00054858	0.511

Masses of the Proton, Neutron, and Electron

Let us look at a few examples of nuclides expressed in the ${}^{A}_{Z}\mathrm{X}_{N}$ notation. The nucleus of the simplest atom, hydrogen, is a single proton, or ${}^{1}_{1}\mathrm{H}$ (the zero for no neutrons is often omitted). To check this symbol, refer to the periodic table—you see that the atomic number Z of hydrogen is 1. Since you are given that there are no neutrons, the mass number A is also 1. Suppose you are told that the helium nucleus or α particle has two protons and two neutrons. You can then see that it is written ${}^{4}_{2}\mathrm{He}_{2}$. There is a scarce form of hydrogen found in nature called deuterium; its nucleus has one proton and one neutron and, hence, twice the mass of common hydrogen. The symbol for deuterium is, thus, ${}^{2}_{1}\mathrm{H}_{1}$ (sometimes D is used, as for deuterated water $\mathrm{D_2O}$). An even rarer—and radioactive—form of hydrogen is called tritium, since it has a single proton and two neutrons, and it is written ${}^{3}_{1}\mathrm{H}_{2}$. These three varieties of hydrogen have nearly identical chemistries, but the nuclei differ greatly in mass, stability, and other characteristics. Nuclei (such as those of hydrogen) having the same Z and different N s are defined to be **isotopes** of the same element.

There is some redundancy in the symbols A, X, Z, and N . If the element X is known, then Z can be found in a periodic table and is always the same for a given element. If both A and X are known, then N can also be determined (first find Z; then, $N=A-Z$). Thus the simpler notation for nuclides is
Equation:

$${}^{A}\mathrm{X},$$

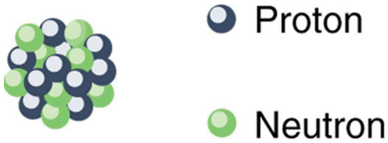
which is sufficient and is most commonly used. For example, in this simpler notation, the three isotopes of hydrogen are ${}^{1}\mathrm{H}$, ${}^{2}\mathrm{H}$, and ${}^{3}\mathrm{H}$, while the α particle is ${}^{4}\mathrm{He}$. We read this backward, saying helium-4 for ${}^{4}\mathrm{He}$, or uranium-238 for ${}^{238}\mathrm{U}$. So for ${}^{238}\mathrm{U}$, should we need to know, we can determine that $Z=92$ for uranium from the periodic table, and, thus, $N=238-92=146$.

A variety of experiments indicate that a nucleus behaves something like a tightly packed ball of nucleons, as illustrated in [link]. These nucleons have large kinetic energies and, thus, move rapidly in very close contact. Nucleons can be separated by a large force, such as in a collision with another nucleus, but resist strongly being pushed closer together. The most compelling evidence that nucleons are closely packed in a nucleus is that the **radius of a nucleus**, r, is found to be given approximately by

Equation:

$$r = r_0 A^{1/3},$$

where $r_0 = 1.2 \text{ fm}$ and A is the mass number of the nucleus. Note that $r^3 \propto A$. Since many nuclei are spherical, and the volume of a sphere is $V = (4/3)\pi r^3$, we see that $V \propto A$ —that is, the volume of a nucleus is proportional to the number of nucleons in it. This is what would happen if you pack nucleons so closely that there is no empty space between them.



A model of the nucleus.

Nucleons are held together by nuclear forces and resist both being pulled apart and pushed inside one another. The volume of the nucleus is the sum of the volumes of the nucleons in it, here shown in different colors to represent protons and neutrons.

Example:
How Small and Dense Is a Nucleus?
(a) Find the radius of an iron-56 nucleus. (b) Find its approximate density in kg/m^3, approximating the mass of ^{56}Fe to be 56 u.
Strategy and Concept

(a) Finding the radius of ^{56}Fe is a straightforward application of $r = r_0 A^{1/3}$, given $A = 56$. (b) To find the approximate density, we assume the nucleus is spherical (this one actually is), calculate its volume using the radius found in part (a), and then find its density from $\rho = m/V$. Finally, we will need to convert density from units of u/fm^3 to kg/m^3.

Solution

(a) The radius of a nucleus is given by

Equation:

$$r = r_0 A^{1/3}.$$

Substituting the values for r_0 and A yields

Equation:

$$\begin{aligned} r &= (1.2\ \text{fm})(56)^{1/3} = (1.2\ \text{fm})(3.83) \\ &= 4.6\ \text{fm}. \end{aligned}$$

(b) Density is defined to be $\rho = m/V$, which for a sphere of radius r is

Equation:

$$\rho = \frac{m}{V} = \frac{m}{(4/3)\pi r^3}.$$

Substituting known values gives

Equation:

$$\begin{aligned} \rho &= \frac{56\ \text{u}}{(1.33)(3.14)(4.6\ \text{fm})^3} \\ &= 0.138\ \text{u/fm}^3. \end{aligned}$$

Converting to units of kg/m^3, we find

Equation:

$$\begin{aligned} \rho &= (0.138\ \text{u/fm}^3)(1.66 \times 10^{-27}\ \text{kg/u})\left(\frac{1\ \text{fm}}{10^{-15}\ \text{m}}\right) \\ &= 2.3 \times 10^{17}\ \text{kg/m}^3. \end{aligned}$$

Discussion

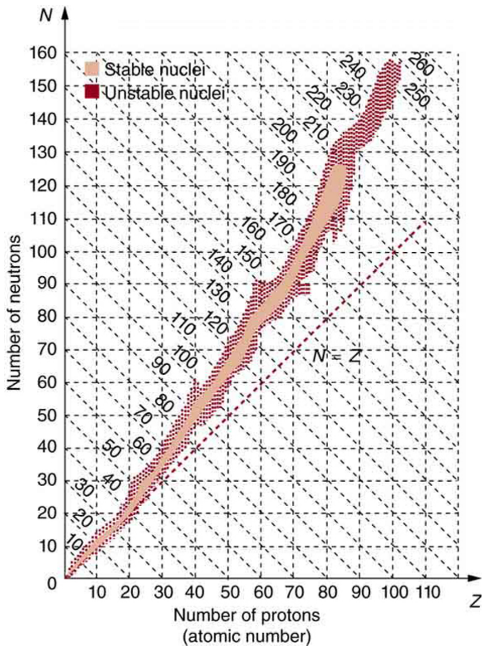
(a) The radius of this medium-sized nucleus is found to be approximately 4.6 fm, and so its diameter is about 10 fm, or 10^{-14} m. In our discussion of Rutherford's discovery of the nucleus, we noticed that it is about 10^{-15} m in diameter (which is for lighter nuclei), consistent with this result to an order of magnitude. The nucleus is much smaller in diameter than the typical atom, which has a diameter of the order of 10^{-10} m.
(b) The density found here is so large as to cause disbelief. It is consistent with earlier discussions we have had about the nucleus being very small and containing nearly all of the mass of the atom. Nuclear densities, such as found here, are about 2×10^{14} times greater than that of water, which has a density of "only" $10^3\ \mathrm{kg/m^3}$. One cubic meter of nuclear matter, such as found in a neutron star, has the same mass as a cube of water 61 km on a side.

Nuclear Forces and Stability

What forces hold a nucleus together? The nucleus is very small and its protons, being positive, exert tremendous repulsive forces on one another. (The Coulomb force increases as charges get closer, since it is proportional to $1/r^2$, even at the tiny distances found in nuclei.) The answer is that two previously unknown forces hold the nucleus together and make it into a tightly packed ball of nucleons. These forces are called the *weak and strong nuclear forces*. Nuclear forces are so short ranged that they fall to zero strength when nucleons are separated by only a few fm. However, like glue, they are strongly attracted when the nucleons get close to one another. The strong nuclear force is about 100 times more attractive than the repulsive EM force, easily holding the nucleons together. Nuclear forces become extremely repulsive if the nucleons get too close, making nucleons strongly resist being pushed inside one another, something like ball bearings.

The fact that nuclear forces are very strong is responsible for the very large energies emitted in nuclear decay. During decay, the forces do work, and since work is force times the distance ($W = Fd \cos \theta$), a large force can result in a large emitted energy. In fact, we know that there are *two* distinct nuclear forces because of the different types of nuclear decay—the strong nuclear force is responsible for α decay, while the weak nuclear force is responsible for β decay.

The many stable and unstable nuclei we have explored, and the hundreds we have not discussed, can be arranged in a table called the **chart of the nuclides**, a simplified version of which is shown in [link]. Nuclides are located on a plot of N versus Z. Examination of a detailed chart of the nuclides reveals patterns in the characteristics of nuclei, such as stability, abundance, and types of decay, analogous to but more complex than the systematics in the periodic table of the elements.



Simplified chart of the nuclides, a graph of N versus Z for known nuclides. The patterns of stable and unstable nuclides reveal characteristics of the nuclear forces. The dashed line is for $N = Z$. Numbers along diagonals are mass numbers A.

In principle, a nucleus can have any combination of protons and neutrons, but [link] shows a definite pattern for those that are stable. For low-mass nuclei, there is a strong tendency for N and Z to be nearly equal. This means that the nuclear force is more attractive when $N = Z$. More detailed examination reveals greater stability when N and Z are even numbers—nuclear forces are more attractive when neutrons and protons are in pairs. For increasingly higher masses, there are progressively more neutrons than protons in stable nuclei. This is due to the ever-growing repulsion between protons. Since nuclear forces are short ranged, and the Coulomb force is long ranged, an excess of neutrons keeps the protons a little farther apart, reducing Coulomb repulsion. Decay modes of nuclides out of the region of stability consistently produce nuclides closer to the region of stability. There are more stable nuclei having certain numbers of protons and neutrons, called **magic numbers**. Magic numbers indicate a shell structure for the nucleus in which closed shells are more stable. Nuclear shell theory has been very successful in explaining nuclear energy levels, nuclear decay, and the greater stability of nuclei with closed shells. We have been producing ever-heavier transuranic elements since the early 1940s, and we have now produced the element with $Z = 118$. There are theoretical predictions of an island of relative stability for nuclei with such high Z s.

The German-born American physicist Maria Goeppert Mayer (1906–1972)

shared the 1963 Nobel Prize in physics with J. Jensen for the creation of the nuclear shell model. This successful nuclear model has nucleons filling shells analogous to electron shells in atoms. It was inspired by patterns observed in nuclear properties. (credit: Nobel Foundation via Wikimedia Commons)

Section Summary

- Two particles, both called nucleons, are found inside nuclei. The two types of nucleons are protons and neutrons; they are very similar, except that the proton is positively charged while the neutron is neutral. Some of their characteristics are given in [link] and compared with those of the electron. A mass unit convenient to atomic and nuclear processes is the unified atomic mass unit (u), defined to be
  Equation:

  $$1\ \text{u} = 1.6605 \times 10^{-27}\ \text{kg} = 931.46\ \text{MeV}/c^2.$$

- A nuclide is a specific combination of protons and neutrons, denoted by
  Equation:

$$ {}^{A}_{Z}\mathrm{X}_{N} \text{ or simply} {}^{A}\mathrm{X}, $$

Z is the number of protons or atomic number, X is the symbol for the element, N is the number of neutrons, and A is the mass number or the total number of protons and neutrons,
Equation:

$$ A = N + Z. $$

- Nuclides having the same Z but different N are isotopes of the same element.
- The radius of a nucleus, r, is approximately
  Equation:

  $$ r = r_0 A^{1/3}, $$

  where $r_0 = 1.2\ \text{fm}$. Nuclear volumes are proportional to A. There are two nuclear forces, the weak and the strong. Systematics in nuclear stability seen on the chart of the nuclides indicate that there are shell closures in nuclei for values of Z and N equal to the magic numbers, which correspond to highly stable nuclei.

Conceptual Questions

Exercise:

Problem:

The weak and strong nuclear forces are basic to the structure of matter. Why we do not experience them directly?

Exercise:

Problem:

Define and make clear distinctions between the terms neutron, nucleon, nucleus, nuclide, and neutrino.

Exercise:

Problem:

What are isotopes? Why do different isotopes of the same element have similar chemistries?

Problems & Exercises

Exercise:

Problem:

Verify that a 2.3×10^{17} kg mass of water at normal density would make a cube 60 km on a side, as claimed in [link]. (This mass at nuclear density would make a cube 1.0 m on a side.)

Solution:
Equation:

$$\begin{aligned} m = \rho V = \rho d^3 \quad \Rightarrow \quad a &= \left(\frac{m}{\rho}\right)^{1/3} = \left(\frac{2.3\times10^{17}\text{ kg}}{1000\text{ kg/m}^3}\right)^{\frac{1}{3}} \\ &= 61 \times 10^3\text{ m} = 61\text{ km} \end{aligned}$$

Exercise:

Problem:

Find the length of a side of a cube having a mass of 1.0 kg and the density of nuclear matter, taking this to be $2.3\times10^{17}\text{ kg/m}^3$.

Exercise:

Problem: What is the radius of an α particle?

Solution:

1.9 fm

Exercise:

Problem:

Find the radius of a ^{238}Pu nucleus. ^{238}Pu is a manufactured nuclide that is used as a power source on some space probes.

Exercise:

Problem:

(a) Calculate the radius of ^{58}Ni, one of the most tightly bound stable nuclei.

(b) What is the ratio of the radius of ^{58}Ni to that of ^{258}Ha, one of the largest nuclei ever made? Note that the radius of the largest nucleus is still much smaller than the size of an atom.

Solution:

(a) 4.6 fm

(b) $0.61\text{ to }1$

Exercise:

Problem:

The unified atomic mass unit is defined to be $1\text{ u}=1.6605\times 10^{-27}\text{ kg}$. Verify that this amount of mass converted to energy yields 931.5 MeV. Note that you must use four-digit or better values for c and $\mid q_e \mid$.

Exercise:

Problem:

What is the ratio of the velocity of a β particle to that of an α particle, if they have the same nonrelativistic kinetic energy?

Solution:

$85.4\text{ to }1$

Exercise:

Problem:

If a 1.50-cm-thick piece of lead can absorb 90.0% of the γ rays from a radioactive source, how many centimeters of lead are needed to absorb all but 0.100% of the γ rays?

Exercise:

Problem:

The detail observable using a probe is limited by its wavelength. Calculate the energy of a γ-ray photon that has a wavelength of 1×10^{-16} m, small enough to detect details about one-tenth the size of a nucleon. Note that a photon having this energy is difficult to produce and interacts poorly with the nucleus, limiting the practicability of this probe.

Solution:

12.4 GeV

Exercise:

Problem:

(a) Show that if you assume the average nucleus is spherical with a radius $r = r_0 A^{1/3}$, and with a mass of A u, then its density is independent of A.

(b) Calculate that density in $\mathrm{u/fm^3}$ and $\mathrm{kg/m^3}$, and compare your results with those found in [link] for $^{56}\mathrm{Fe}$.

Exercise:

Problem:

What is the ratio of the velocity of a 5.00-MeV β ray to that of an α particle with the same kinetic energy? This should confirm that βs travel much faster than αs even when relativity is taken into consideration. (See also [link].)

Solution:

19.3 to 1

Exercise:

Problem:

(a) What is the kinetic energy in MeV of a β ray that is traveling at $0.998c$? This gives some idea of how energetic a β ray must be to travel at nearly the same speed as a γ ray. (b) What is the velocity of the γ ray relative to the β ray?

Glossary

atomic mass
: the total mass of the protons, neutrons, and electrons in a single atom

atomic number
: number of protons in a nucleus

chart of the nuclides
: a table comprising stable and unstable nuclei

isotopes
: nuclei having the same Z and different Ns

magic numbers
: a number that indicates a shell structure for the nucleus in which closed shells are more stable

mass number
: number of nucleons in a nucleus

neutron
: a neutral particle that is found in a nucleus

nucleons
: the particles found inside nuclei

nucleus
: a region consisting of protons and neutrons at the center of an atom

nuclide

a type of atom whose nucleus has specific numbers of protons and neutrons

protons
: the positively charged nucleons found in a nucleus

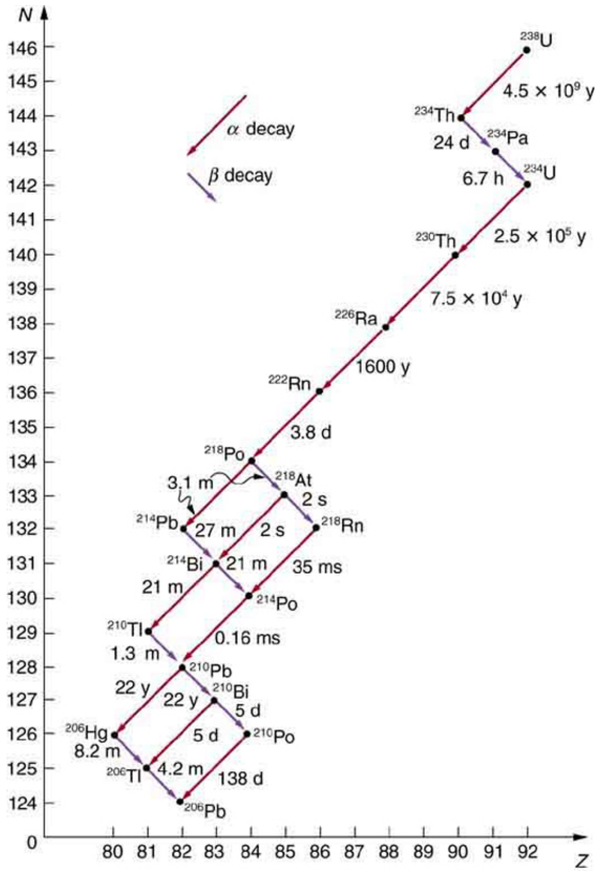
radius of a nucleus
: the radius of a nucleus is $r = r_0 A^{1/3}$

Nuclear Decay and Conservation Laws

- Define and discuss nuclear decay.
- State the conservation laws.
- Explain parent and daughter nucleus.
- Calculate the energy emitted during nuclear decay.

Nuclear **decay** has provided an amazing window into the realm of the very small. Nuclear decay gave the first indication of the connection between mass and energy, and it revealed the existence of two of the four basic forces in nature. In this section, we explore the major modes of nuclear decay; and, like those who first explored them, we will discover evidence of previously unknown particles and conservation laws.

Some nuclides are stable, apparently living forever. Unstable nuclides decay (that is, they are radioactive), eventually producing a stable nuclide after many decays. We call the original nuclide the **parent** and its decay products the **daughters**. Some radioactive nuclides decay in a single step to a stable nucleus. For example, ^{60}Co is unstable and decays directly to ^{60}Ni, which is stable. Others, such as ^{238}U, decay to another unstable nuclide, resulting in a **decay series** in which each subsequent nuclide decays until a stable nuclide is finally produced. The decay series that starts from ^{238}U is of particular interest, since it produces the radioactive isotopes ^{226}Ra and ^{210}Po, which the Curies first discovered (see [link]). Radon gas is also produced (^{222}Rn in the series), an increasingly recognized naturally occurring hazard. Since radon is a noble gas, it emanates from materials, such as soil, containing even trace amounts of ^{238}U and can be inhaled. The decay of radon and its daughters produces internal damage. The ^{238}U decay series ends with ^{206}Pb, a stable isotope of lead.



The decay series produced by ^{238}U, the most common uranium isotope. Nuclides are graphed in the same manner as in the chart of nuclides. The type of decay for each member of the series is shown, as well as the half-lives. Note that some nuclides decay by more than one mode. You can see why radium and polonium are found in uranium ore. A stable isotope of lead is the end product of the series.

Note that the daughters of α decay shown in [link] always have two fewer protons and two fewer neutrons than the parent. This seems reasonable, since we know that α decay is the emission of a ^4He nucleus, which has two protons and two neutrons. The daughters of β decay have one less neutron and one more proton than their parent. Beta decay is a little more subtle, as we shall see. No γ decays are shown in the figure, because they do not produce a daughter that differs from the parent.

Alpha Decay

In **alpha decay**, a ^{4}He nucleus simply breaks away from the parent nucleus, leaving a daughter with two fewer protons and two fewer neutrons than the parent (see [link]). One example of α decay is shown in [link] for ^{238}U. Another nuclide that undergoes α decay is ^{239}Pu. The decay equations for these two nuclides are

Equation:

$$^{238}\text{U} \rightarrow {}^{234}\text{Th}_{92}^{234} + {}^{4}\text{He}$$

and

Equation:

$$^{239}\text{Pu} \rightarrow {}^{235}\text{U} + {}^{4}\text{He}.$$

Alpha decay is the separation of a ^{4}He nucleus from the parent. The daughter nucleus has two fewer protons and two fewer neutrons than the parent. Alpha decay occurs spontaneously only if the daughter and ^{4}He nucleus have less total mass than the parent.

If you examine the periodic table of the elements, you will find that Th has $Z = 90$, two fewer than U, which has $Z = 92$. Similarly, in the second **decay equation**, we see that U has two fewer protons than Pu, which has $Z = 94$. The general rule for α decay is best written in the format $^{A}_{Z}\text{X}_{N}$. If a certain nuclide is known to α decay (generally this information must be looked up in a table of isotopes, such as in Appendix B), its α **decay equation** is

Equation:

$$ {}_{Z}^{A}\text{X}_{N} \rightarrow {}_{Z-2}^{A-4}\text{Y}_{N-2} + {}_{2}^{4}\text{He}_{2} \quad (\alpha \text{ decay}) $$

where Y is the nuclide that has two fewer protons than X, such as Th having two fewer than U. So if you were told that ^{239}Pu α decays and were asked to write the complete decay equation, you would first look up which element has two fewer protons (an atomic number two lower) and find that this is uranium. Then since four nucleons have broken away from the original 239, its atomic mass would be 235.

It is instructive to examine conservation laws related to α decay. You can see from the equation ${}_{Z}^{A}\text{X}_{N} \rightarrow {}_{Z-2}^{A-4}\text{Y}_{N-2} + {}_{2}^{4}\text{He}_{2}$ that total charge is conserved. Linear and angular momentum are conserved, too. Although conserved angular momentum is not of great consequence in this type of decay, conservation of linear momentum has interesting consequences. If the nucleus is at rest when it decays, its momentum is zero. In that case, the fragments must fly in opposite directions with equal-magnitude momenta so that total momentum remains zero. This results in the α particle carrying away most of the energy, as a bullet from a heavy rifle carries away most of the energy of the powder burned to shoot it. Total mass–energy is also conserved: the energy produced in the decay comes from conversion of a fraction of the original mass. As discussed in Atomic Physics, the general relationship is

Equation:

$$ E = (\Delta m)c^2. $$

Here, E is the **nuclear reaction energy** (the reaction can be nuclear decay or any other reaction), and Δm is the difference in mass between initial and final products. When the final products have less total mass, Δm is positive, and the reaction releases energy (is exothermic). When the products have greater total mass, the reaction is endothermic (Δm is negative) and must be induced with an energy input. For α decay to be spontaneous, the decay products must have smaller mass than the parent.

Example:

Alpha Decay Energy Found from Nuclear Masses

Find the energy emitted in the α decay of ^{239}Pu.

Strategy

Nuclear reaction energy, such as released in α decay, can be found using the equation $E = (\Delta m)c^2$. We must first find Δm, the difference in mass between the parent nucleus and the products of the decay. This is easily done using masses given in Appendix A.

Solution

The decay equation was given earlier for ^{239}Pu ; it is

Equation:

$$^{239}\text{Pu} \rightarrow {}^{235}\text{U} + {}^{4}\text{He}.$$

Thus the pertinent masses are those of ^{239}Pu, ^{235}U, and the α particle or ^{4}He, all of which are listed in Appendix A. The initial mass was $m(^{239}\text{Pu}) = 239.052157\ \text{u}$. The final mass is the sum $m(^{235}\text{U})$+$m(^{4}\text{He})$= $235.043924\ \text{u} + 4.002602\ \text{u} = 239.046526\ \text{u}$. Thus,

Equation:

$$\begin{aligned} \Delta m &= m(^{239}\text{Pu}) - [m(^{235}\text{U}) + m(^{4}\text{He})] \\ &= 239.052157\ \text{u} - 239.046526\ \text{u} \\ &= 0.0005631\ \text{u}. \end{aligned}$$

Now we can find E by entering Δm into the equation:

Equation:

$$E = (\Delta m)c^2 = (0.005631\ \text{u})c^2.$$

We know $1\ \text{u} = 931.5\ \text{MeV}/c^2$, and so

Equation:

$$E = (0.005631)(931.5\ \text{MeV}/c^2)(c^2) = 5.25\ \text{MeV}.$$

Discussion

The energy released in this α decay is in the MeV range, about 10^6 times as great as typical chemical reaction energies, consistent with many previous discussions. Most of this energy becomes kinetic energy of the α particle (or ^{4}He nucleus), which moves away at high speed. The energy carried away by the recoil of the ^{235}U nucleus is much smaller in order to conserve momentum. The ^{235}U nucleus can be left in an excited state to later emit photons (γ rays). This decay is spontaneous and releases energy, because the products have less mass than the parent nucleus. The question of why the products have less mass will be discussed in Binding Energy. Note that the masses given in Appendix A are atomic masses of neutral atoms, including their electrons. The mass of the electrons is the same before and after α decay, and so their masses subtract out when finding Δm. In this case, there are 94 electrons before and after the decay.

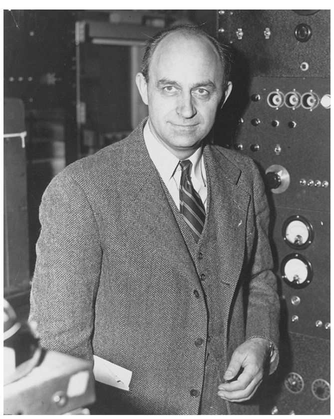
Beta Decay

There are actually *three* types of **beta decay**. The first discovered was “ordinary” beta decay and is called β^- decay or electron emission. The symbol β^- represents *an electron emitted in nuclear beta decay*. Cobalt-60 is a nuclide that β^- decays in the following manner:

Equation:

$$^{60}\text{Co} \rightarrow {}^{60}\text{Ni} + \beta^- + \text{neutrino}.$$

The **neutrino** is a particle emitted in beta decay that was unanticipated and is of fundamental importance. The neutrino was not even proposed in theory until more than 20 years after beta decay was known to involve electron emissions. Neutrinos are so difficult to detect that the first direct evidence of them was not obtained until 1953. Neutrinos are nearly massless, have no charge, and do not interact with nucleons via the strong nuclear force. Traveling approximately at the speed of light, they have little time to affect any nucleus they encounter. This is, owing to the fact that they have no charge (and they are not EM waves), they do not interact through the EM force. They do interact via the relatively weak and very short range weak nuclear force. Consequently, neutrinos escape almost any detector and penetrate almost any shielding. However, neutrinos do carry energy, angular momentum (they are fermions with half-integral spin), and linear momentum away from a beta decay. When accurate measurements of beta decay were made, it became apparent that energy, angular momentum, and linear momentum were not accounted for by the daughter nucleus and electron alone. Either a previously unsuspected particle was carrying them away, or three conservation laws were being violated. Wolfgang Pauli made a formal proposal for the existence of neutrinos in 1930. The Italian-born American physicist Enrico Fermi (1901–1954) gave neutrinos their name, meaning little neutral ones, when he developed a sophisticated theory of beta decay (see [link]). Part of Fermi's theory was the identification of the weak nuclear force as being distinct from the strong nuclear force and in fact responsible for beta decay.



Enrico Fermi was nearly unique among 20th-century physicists—he made significant contributions both as an experimentalist and a theorist. His many contributions to theoretical

physics included the identification of the weak nuclear force. The fermi (fm) is named after him, as are an entire class of subatomic particles (fermions), an element (Fermium), and a major research laboratory (Fermilab). His experimental work included studies of radioactivity, for which he won the 1938 Nobel Prize in physics, and creation of the first nuclear chain reaction. (credit: United States Department of Energy, Office of Public Affairs)

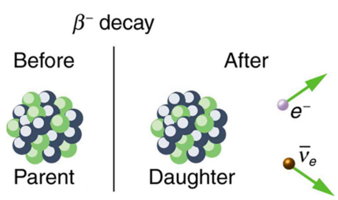
The neutrino also reveals a new conservation law. There are various families of particles, one of which is the electron family. We propose that the number of members of the electron family is constant in any process or any closed system. In our example of beta decay, there are no members of the electron family present before the decay, but after, there is an electron and a neutrino. So electrons are given an electron family number of $+1$. The neutrino in β^{-} decay is an **electron's antineutrino**, given the symbol ν_e, where ν is the Greek letter nu, and the subscript e means this neutrino is related to the electron. The bar indicates this is a particle of **antimatter**. (All particles have antimatter counterparts that are nearly identical except that they have the opposite charge. Antimatter is almost entirely absent on Earth, but it is found in nuclear decay and other nuclear and particle reactions as well as in outer space.) The electron's antineutrino ν_e, being antimatter, has an electron family number of –1. The total is zero, before and after the decay. The new conservation law, obeyed in all circumstances, states that the *total electron family number is constant*. An electron cannot be created without also creating an antimatter family member. This law is analogous to the conservation of charge in a situation where total charge is originally zero, and equal amounts of positive and negative charge must be created in a reaction to keep the total zero.

If a nuclide ${}^{A}_{Z}\text{X}_{N}$ is known to β^- decay, then its β^- decay equation is
Equation:

$${}^{A}_{Z}\text{X}_{N} \rightarrow {}_{Z+1}^{A}\text{Y}_{N-1} + \beta^- + \bar{\nu}_e \ (\beta^- \text{ decay}),$$

where Y is the nuclide having one more proton than X (see [link]). So if you know that a certain nuclide β^- decays, you can find the daughter nucleus by first looking up Z for the parent and then determining which element has atomic number $Z+1$. In the example of the β^- decay of ${}^{60}\text{Co}$ given earlier, we see that $Z=27$ for Co and $Z=28$ is Ni. It is as if one of the neutrons in the parent nucleus decays into a proton, electron, and neutrino. In fact, neutrons outside of nuclei do just that—they live only an average of a few minutes and β^- decay in the following manner:
Equation:

$$\text{n} \rightarrow \text{p} + \beta^- + \bar{\nu}_e.$$



In β^- decay, the parent nucleus emits an electron and an antineutrino. The daughter nucleus has one more proton and one less neutron than its parent. Neutrinos interact so weakly that they are almost never directly observed, but they play a fundamental role in particle physics.

We see that charge is conserved in β^- decay, since the total charge is Z before and after the decay. For example, in ^{60}Co decay, total charge is 27 before decay, since cobalt has $Z = 27$. After decay, the daughter nucleus is Ni, which has $Z = 28$, and there is an electron, so that the total charge is also $28 + (-1)$ or 27. Angular momentum is conserved, but not obviously (you have to examine the spins and angular momenta of the final products in detail to verify this). Linear momentum is also conserved, again imparting most of the decay energy to the electron and the antineutrino, since they are of low and zero mass, respectively. Another new conservation law is obeyed here and elsewhere in nature. *The total number of nucleons A is conserved*. In ^{60}Co decay, for example, there are 60 nucleons before and after the decay. Note that total A is also conserved in α decay. Also note that the total number of protons changes, as does the total number of neutrons, so that total Z and total N are *not* conserved in β^- decay, as they are in α decay. Energy released in β^- decay can be calculated given the masses of the parent and products.

Example:

β^- Decay Energy from Masses

Find the energy emitted in the β^- decay of ^{60}Co.

Strategy and Concept

As in the preceding example, we must first find Δm, the difference in mass between the parent nucleus and the products of the decay, using masses given in Appendix A. Then the emitted energy is calculated as before, using $E = (\Delta m)c^2$. The initial mass is just that of the parent nucleus, and the final mass is that of the daughter nucleus and the electron created in the decay. The neutrino is massless, or nearly so. However, since the masses given in Appendix A are for neutral atoms, the daughter nucleus has one more electron than the parent, and so the extra electron mass that corresponds to the β^- is included in the atomic mass of Ni. Thus,

Equation:

$$\Delta m = m(^{60}\text{Co}) - m(^{60}\text{Ni}\ ).$$

Solution

The β^- decay equation for ^{60}Co is

Equation:

$${}^{60}_{27}\text{Co}_{33} \to {}^{60}_{28}\text{Ni}_{32} + \beta^- + \nu_e.$$

As noticed,

Equation:

$$\Delta m = m(^{60}\text{Co}) - m(^{60}\text{Ni}\ ).$$

Entering the masses found in Appendix A gives

Equation:

$$\Delta m = 59.933820 \text{ u} - 59.930789 \text{ u} = 0.003031 \text{ u}.$$

Thus,

Equation:

$$E = (\Delta m)c^2 = (0.003031 \text{ u})c^2.$$

Using $1 \text{ u} = 931.5 \text{ MeV}/c^2$, we obtain

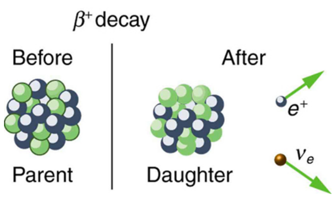
Equation:

$$E = (0.003031)(931.5 \text{ MeV}/c^2)(c^2) = 2.82 \text{ MeV}.$$

Discussion and Implications

Perhaps the most difficult thing about this example is convincing yourself that the β^- mass is included in the atomic mass of ^{60}Ni. Beyond that are other implications. Again the decay energy is in the MeV range. This energy is shared by all of the products of the decay. In many ^{60}Co decays, the daughter nucleus ^{60}Ni is left in an excited state and emits photons (γ rays). Most of the remaining energy goes to the electron and neutrino, since the recoil kinetic energy of the daughter nucleus is small. One final note: the electron emitted in β^- decay is created in the nucleus at the time of decay.

The second type of beta decay is less common than the first. It is β^+decay. Certain nuclides decay by the emission of a *positive* electron. This is **antielectron** or **positron decay** (see [link]).



β^+ decay is the emission of a positron that eventually finds an electron to annihilate, characteristically producing gammas in opposite directions.

The antielectron is often represented by the symbol e^+, but in beta decay it is written as β^+ to indicate the antielectron was emitted in a nuclear decay. Antielectrons are the antimatter counterpart to electrons, being nearly identical, having the same mass, spin, and so on, but having a positive charge and an electron family number of –1. When a **positron** encounters an electron, there is a mutual annihilation in which all the mass of the antielectron-electron pair is converted into pure photon energy. (The reaction, $e^+ + e^- \rightarrow \gamma + \gamma$, conserves electron family number as well as all other conserved quantities.) If a nuclide ${}^{A}_{Z}\mathrm{X}_N$ is known to β^+ decay, then its β^+**decay equation** is

Equation:

$${}^{A}_{Z}\mathrm{X}_N \rightarrow {}_{Z-1}^{A}\mathrm{Y}_{N+1} + \beta^+ + \nu_e \ (\beta^+ \text{ decay}),$$

where Y is the nuclide having one less proton than X (to conserve charge) and ν_e is the symbol for the **electron's neutrino**, which has an electron family number of $+1$. Since an antimatter member of the electron family (the β^+) is created in the decay, a matter member of the family (here the ν_e) must also be created. Given, for example, that ${}^{22}\mathrm{Na}$ β^+ decays, you can write its full decay equation by first finding that $Z = 11$ for ${}^{22}\mathrm{Na}$, so that the daughter nuclide will have $Z = 10$, the atomic number for neon. Thus the β^+ decay equation for ${}^{22}\mathrm{Na}$ is

Equation:

$${}^{22}_{11}\mathrm{Na}_{11} \rightarrow {}^{22}_{10}\mathrm{Ne}_{12} + \beta^+ + \nu_e.$$

In β^+ decay, it is as if one of the protons in the parent nucleus decays into a neutron, a positron, and a neutrino. Protons do not do this outside of the nucleus, and so the decay is due to the complexities of the nuclear force. Note again that the total number of nucleons is constant in this and any other reaction. To find the energy emitted in β^+ decay, you must again count the number of electrons in the neutral atoms, since atomic masses are used. The daughter has one less electron than the parent, and one electron mass is created in the decay. Thus, in β^+ decay,

Equation:

$$\Delta m = m(\text{parent}) - [m(\text{daughter}) + 2m_e],$$

since we use the masses of neutral atoms.

Electron capture is the third type of beta decay. Here, a nucleus captures an inner-shell electron and undergoes a nuclear reaction that has the same effect as β^+ decay. Electron capture is sometimes denoted by the letters EC. We know that electrons cannot reside in the nucleus, but this is a nuclear reaction that consumes the electron and occurs spontaneously only when the products have less mass than the parent plus the electron. If a nuclide ${}^{A}_{Z}\mathrm{X}_N$ is known to undergo electron capture, then its **electron capture equation** is

Equation:

$$ {}_{Z}^{A}\text{X}_{N} + e^{-} \rightarrow {}_{Z-1}^{A}\text{Y}_{N+1} + \nu_e (\text{electron capture, or EC}). $$

Any nuclide that can β^+ decay can also undergo electron capture (and often does both). The same conservation laws are obeyed for EC as for β^+ decay. It is good practice to confirm these for yourself.

All forms of beta decay occur because the parent nuclide is unstable and lies outside the region of stability in the chart of nuclides. Those nuclides that have relatively more neutrons than those in the region of stability will β^- decay to produce a daughter with fewer neutrons, producing a daughter nearer the region of stability. Similarly, those nuclides having relatively more protons than those in the region of stability will β^- decay or undergo electron capture to produce a daughter with fewer protons, nearer the region of stability.

Gamma Decay

Gamma decay is the simplest form of nuclear decay—it is the emission of energetic photons by nuclei left in an excited state by some earlier process. Protons and neutrons in an excited nucleus are in higher orbitals, and they fall to lower levels by photon emission (analogous to electrons in excited atoms). Nuclear excited states have lifetimes typically of only about 10^{-14} s, an indication of the great strength of the forces pulling the nucleons to lower states. The γ decay equation is simply

Equation:

$$ {}_{Z}^{A}\text{X}_{N}^{*} \rightarrow {}_{Z}^{A}\text{X}_{N} + \gamma_1 + \gamma_2 + \cdots \ (\gamma \text{ decay}) $$

where the asterisk indicates the nucleus is in an excited state. There may be one or more γ s emitted, depending on how the nuclide de-excites. In radioactive decay, γ emission is common and is preceded by γ or β decay. For example, when ^{60}Co β^- decays, it most often leaves the daughter nucleus in an excited state, written ^{60}Ni*. Then the nickel nucleus quickly γ decays by the emission of two penetrating γ s:

Equation:

$$ ^{60}\text{Ni}^{*} \rightarrow {}^{60}\text{Ni} + \gamma_1 + \gamma_2. $$

These are called cobalt γ rays, although they come from nickel—they are used for cancer therapy, for example. It is again constructive to verify the conservation laws for gamma decay. Finally, since γ decay does not change the nuclide to another species, it is not prominently featured in charts of decay series, such as that in [link].

There are other types of nuclear decay, but they occur less commonly than α, β, and γ decay. Spontaneous fission is the most important of the other forms of nuclear decay because of its applications in nuclear power and weapons. It is covered in the next chapter.

Section Summary

- When a parent nucleus decays, it produces a daughter nucleus following rules and conservation laws. There are three major types of nuclear decay, called alpha (α), beta (β), and gamma (γ). The α decay equation is
  Equation:
  $$^{A}_{Z}\mathrm{X}_{N} \rightarrow {}^{A-4}_{Z-2}\mathrm{Y}_{N-2} + {}^{4}_{2}\mathrm{He}_{2}.$$
- Nuclear decay releases an amount of energy E related to the mass destroyed Δm by
  Equation:
  $$E = (\Delta m)c^{2}.$$
- There are three forms of beta decay. The β^{-}decay equation is
  Equation:
  $$^{A}_{Z}\mathrm{X}_{N} \rightarrow {}^{A}_{Z+1}\mathrm{Y}_{N-1} + \beta^{-} + \nu_{e}.$$
- The β^{+} decay equation is
  Equation:
  $$^{A}_{Z}\mathrm{X}_{N} \rightarrow {}^{A}_{Z-1}\mathrm{Y}_{N+1} + \beta^{+} + \nu_{e}.$$
- The electron capture equation is
  Equation:
  $$^{A}_{Z}\mathrm{X}_{N} + e^{-} \rightarrow {}^{A}_{Z-1}\mathrm{Y}_{N+1} + \nu_{e}.$$
- β^{-} is an electron, β^{+} is an antielectron or positron, ν_{e} represents an electron's neutrino, and ν_{e} is an electron's antineutrino. In addition to all previously known conservation laws, two new ones arise— conservation of electron family number and conservation of the total number of nucleons. The γ decay equation is
  Equation:
  $$^{A}_{Z}\mathrm{X}^{*}_{N} \rightarrow {}^{A}_{Z}\mathrm{X}_{N} + \gamma_{1} + \gamma_{2} + \cdots$$
  γ is a high-energy photon originating in a nucleus.

Conceptual Questions

Exercise:

Problem:

Star Trek fans have often heard the term “antimatter drive.” Describe how you could use a magnetic field to trap antimatter, such as produced by nuclear decay, and later combine it with matter to produce energy. Be specific about the type of antimatter, the need for vacuum storage, and the fraction of matter converted into energy.

Exercise:

Problem:

What conservation law requires an electron’s neutrino to be produced in electron capture? Note that the electron no longer exists after it is captured by the nucleus.

Exercise:

Problem:

Neutrinos are experimentally determined to have an extremely small mass. Huge numbers of neutrinos are created in a supernova at the same time as massive amounts of light are first produced. When the 1987A supernova occurred in the Large Magellanic Cloud, visible primarily in the Southern Hemisphere and some 100,000 light-years away from Earth, neutrinos from the explosion were observed at about the same time as the light from the blast. How could the relative arrival times of neutrinos and light be used to place limits on the mass of neutrinos?

Exercise:

Problem:

What do the three types of beta decay have in common that is distinctly different from alpha decay?

Problems & Exercises

In the following eight problems, write the complete decay equation for the given nuclide in the complete ${}^{A}_{Z}\text{X}_{N}$ notation. Refer to the periodic table for values of Z.

Exercise:

Problem:

β^- decay of ${}^{3}\text{H}$ (tritium), a manufactured isotope of hydrogen used in some digital watch displays, and manufactured primarily for use in hydrogen bombs.

Solution:

Equation:

$${}^{3}_{1}\text{H}_2 \rightarrow {}^{3}_{2}\text{He}_1 + \beta^- + \nu_e$$

Exercise:

Problem:

β^- decay of ^{40}K, a naturally occurring rare isotope of potassium responsible for some of our exposure to background radiation.

Exercise:

Problem: β^+ decay of ^{50}Mn.

Solution:
Equation:

$$^{50}_{25}M_{25} \rightarrow {}^{50}_{24}\mathrm{Cr}_{26} + \beta^+ + \nu_e$$

Exercise:

Problem: β^+ decay of ^{52}Fe.

Exercise:

Problem: Electron capture by ^{7}Be.

Solution:
Equation:

$$^{7}_{4}\mathrm{Be}_3 + e^- \rightarrow {}^{7}_{3}\mathrm{Li}_4 + \nu_e$$

Exercise:

Problem: Electron capture by ^{106}In.

Exercise:

Problem:

α decay of ^{210}Po, the isotope of polonium in the decay series of ^{238}U that was discovered by the Curies. A favorite isotope in physics labs, since it has a short half-life and decays to a stable nuclide.

Solution:
Equation:

$$^{210}_{84}\mathrm{Po}_{126} \rightarrow {}^{206}_{82}\mathrm{Pb}_{124} + {}^{4}_{2}\mathrm{He}_2$$

Exercise:

Problem:

α decay of $^{226}\mathrm{Ra}$, another isotope in the decay series of $^{238}\mathrm{U}$, first recognized as a new element by the Curies. Poses special problems because its daughter is a radioactive noble gas.

In the following four problems, identify the parent nuclide and write the complete decay equation in the $^{A}_{Z}\mathrm{X}_{N}$ notation. Refer to the periodic table for values of Z.

Exercise:

Problem:

β^- decay producing $^{137}\mathrm{Ba}$. The parent nuclide is a major waste product of reactors and has chemistry similar to potassium and sodium, resulting in its concentration in your cells if ingested.

Solution:
Equation:

$$^{137}_{55}\mathrm{Cs}_{82} \rightarrow {}^{137}_{56}\mathrm{Ba}_{81} + \beta^- + \nu_e$$

Exercise:

Problem:

β^- decay producing $^{90}\mathrm{Y}$. The parent nuclide is a major waste product of reactors and has chemistry similar to calcium, so that it is concentrated in bones if ingested ($^{90}\mathrm{Y}$ is also radioactive.)

Exercise:

Problem:

α decay producing $^{228}\mathrm{Ra}$. The parent nuclide is nearly 100% of the natural element and is found in gas lantern mantles and in metal alloys used in jets ($^{228}\mathrm{Ra}$ is also radioactive).

Solution:
Equation:

$$^{232}_{90}\mathrm{Th}_{142} \rightarrow {}^{228}_{88}\mathrm{Ra}_{140} + {}^{4}_{2}\mathrm{He}_{2}$$

Exercise:

Problem:

α decay producing $^{208}\mathrm{Pb}$. The parent nuclide is in the decay series produced by $^{232}\mathrm{Th}$, the only naturally occurring isotope of thorium.

Exercise:

Problem:

When an electron and positron annihilate, both their masses are destroyed, creating two equal energy photons to preserve momentum. (a) Confirm that the annihilation equation $e^{+} + e^{-} \rightarrow \gamma + \gamma$ conserves charge, electron family number, and total number of nucleons. To do this, identify the values of each before and after the annihilation. (b) Find the energy of each γ ray, assuming the electron and positron are initially nearly at rest. (c) Explain why the two γ rays travel in exactly opposite directions if the center of mass of the electron-positron system is initially at rest.

Solution:

(a)
$\text{charge:}(+1) + (-1) = 0\text{;}\ \ \text{electron family number: } (+1) + (-1) = 0\text{;}\ \ A\text{: } 0 + 0 = 0$

(b) 0.511 MeV

(c) The two γ rays must travel in exactly opposite directions in order to conserve momentum, since initially there is zero momentum if the center of mass is initially at rest.

Exercise:

Problem:

Confirm that charge, electron family number, and the total number of nucleons are all conserved by the rule for α decay given in the equation ${}_{Z}^{A}\mathrm{X}_{N} \rightarrow {}_{Z-2}^{A-4}\mathrm{Y}_{N-2} + {}_{2}^{4}\mathrm{He}_{2}$. To do this, identify the values of each before and after the decay.

Exercise:

Problem:

Confirm that charge, electron family number, and the total number of nucleons are all conserved by the rule for β^{-} decay given in the equation ${}_{Z}^{A}\mathrm{X}_{N} \rightarrow {}_{Z+1}^{A}\mathrm{Y}_{N-1} + \beta^{-} + \nu_{e}$. To do this, identify the values of each before and after the decay.

Solution:
Equation:

$$Z = (Z + 1) - 1;\quad A = A;\quad \text{efn} : 0 = (+1) + (-1)$$

Exercise:

Problem:

Confirm that charge, electron family number, and the total number of nucleons are all conserved by the rule for β^- decay given in the equation ${}^{A}_{Z}\mathrm{X}_N \rightarrow {}^{A}_{Z-1}\mathrm{Y}_{N-1} + \beta^- + \nu_e$. To do this, identify the values of each before and after the decay.

Exercise:

Problem:

Confirm that charge, electron family number, and the total number of nucleons are all conserved by the rule for electron capture given in the equation ${}^{A}_{Z}\mathrm{X}_N + e^- \rightarrow {}^{A}_{Z-1}\mathrm{Y}_{N+1} + \nu_e$. To do this, identify the values of each before and after the capture.

Solution:
Equation:

$$Z - 1 = Z - 1; \quad A = A; \quad \text{efn :}(+1) = (+1)$$

Exercise:

Problem:

A rare decay mode has been observed in which ${}^{222}\mathrm{Ra}$ emits a ${}^{14}\mathrm{C}$ nucleus. (a) The decay equation is ${}^{222}\mathrm{Ra} \rightarrow {}^{A}\mathrm{X} + {}^{14}\mathrm{C}$. Identify the nuclide ${}^{A}\mathrm{X}$. (b) Find the energy emitted in the decay. The mass of ${}^{222}\mathrm{Ra}$ is 222.015353 u.

Exercise:

Problem: (a) Write the complete α decay equation for ${}^{226}\mathrm{Ra}$.

(b) Find the energy released in the decay.

Solution:

(a) ${}^{226}_{88}\mathrm{Ra}_{138} \rightarrow {}^{222}_{86}\mathrm{Rn}_{136} + {}^{4}_{2}\mathrm{He}_2$

(b) 4.87 MeV

Exercise:

Problem: (a) Write the complete α decay equation for ${}^{249}\mathrm{Cf}$.

(b) Find the energy released in the decay.

Exercise:

Problem:

(a) Write the complete β^- decay equation for the neutron. (b) Find the energy released in the decay.

Solution:

(a) $\mathrm{n} \rightarrow \mathrm{p} + \beta^- + \nu_e$

(b)) 0.783 MeV

Exercise:

Problem:

(a) Write the complete β^- decay equation for $^{90}\mathrm{Sr}$, a major waste product of nuclear reactors. (b) Find the energy released in the decay.

Exercise:

Problem:

Calculate the energy released in the β^+ decay of $^{22}\mathrm{Na}$, the equation for which is given in the text. The masses of $^{22}\mathrm{Na}$ and $^{22}\mathrm{Ne}$ are 21.994434 and 21.991383 u, respectively.

Solution:

1.82 MeV

Exercise:

Problem: (a) Write the complete β^+ decay equation for $^{11}\mathrm{C}$.

(b) Calculate the energy released in the decay. The masses of $^{11}\mathrm{C}$ and $^{11}\mathrm{B}$ are 11.011433 and 11.009305 u, respectively.

Exercise:

Problem: (a) Calculate the energy released in the α decay of $^{238}\mathrm{U}$.

(b) What fraction of the mass of a single $^{238}\mathrm{U}$ is destroyed in the decay? The mass of $^{234}\mathrm{Th}$ is 234.043593 u.

(c) Although the fractional mass loss is large for a single nucleus, it is difficult to observe for an entire macroscopic sample of uranium. Why is this?

Solution:

(a) 4.274 MeV

(b) 1.927×10^{-5}

(c) Since U-238 is a slowly decaying substance, only a very small number of nuclei decay on human timescales; therefore, although those nuclei that decay lose a noticeable fraction of their mass, the change in the total mass of the sample is not detectable for a macroscopic sample.

Exercise:

Problem: (a) Write the complete reaction equation for electron capture by ^{7}Be.

(b) Calculate the energy released.

Exercise:

Problem: (a) Write the complete reaction equation for electron capture by ^{15}O.

(b) Calculate the energy released.

Solution:

(a) $^{15}_{8}\text{O}_7 + e^- \rightarrow {}^{15}_{7}\text{N}_8 + \nu_e$

(b) 2.754 MeV

Glossary

parent
: the original state of nucleus before decay

daughter
: the nucleus obtained when parent nucleus decays and produces another nucleus following the rules and the conservation laws

positron
: the particle that results from positive beta decay; also known as an antielectron

decay
: the process by which an atomic nucleus of an unstable atom loses mass and energy by emitting ionizing particles

alpha decay
: type of radioactive decay in which an atomic nucleus emits an alpha particle

beta decay
: type of radioactive decay in which an atomic nucleus emits a beta particle

gamma decay
: type of radioactive decay in which an atomic nucleus emits a gamma particle

decay equation
: the equation to find out how much of a radioactive material is left after a given period of time

nuclear reaction energy
: the energy created in a nuclear reaction

neutrino
: an electrically neutral, weakly interacting elementary subatomic particle

electron’s antineutrino
: antiparticle of electron’s neutrino

positron decay
: type of beta decay in which a proton is converted to a neutron, releasing a positron and a neutrino

antielectron
: another term for positron

decay series
: process whereby subsequent nuclides decay until a stable nuclide is produced

electron’s neutrino
: a subatomic elementary particle which has no net electric charge

antimatter
: composed of antiparticles

electron capture
: the process in which a proton-rich nuclide absorbs an inner atomic electron and simultaneously emits a neutrino

electron capture equation
: equation representing the electron capture

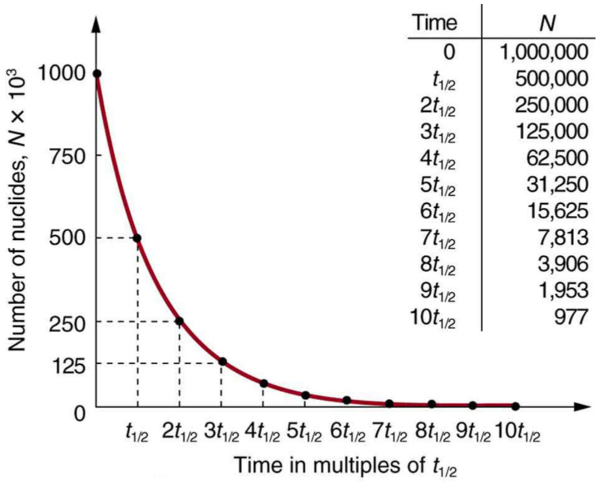
Half-Life and Activity

- Define half-life.
- Define dating.
- Calculate age of old objects by radioactive dating.

Unstable nuclei decay. However, some nuclides decay faster than others. For example, radium and polonium, discovered by the Curies, decay faster than uranium. This means they have shorter lifetimes, producing a greater rate of decay. In this section we explore half-life and activity, the quantitative terms for lifetime and rate of decay.

Half-Life

Why use a term like half-life rather than lifetime? The answer can be found by examining [link], which shows how the number of radioactive nuclei in a sample decreases with time. The *time in which half of the original number of nuclei decay* is defined as the **half-life**, $t_{1/2}$. Half of the remaining nuclei decay in the next half-life. Further, half of that amount decays in the following half-life. Therefore, the number of radioactive nuclei decreases from N to $N/2$ in one half-life, then to $N/4$ in the next, and to $N/8$ in the next, and so on. If N is a large number, then *many* half-lives (not just two) pass before all of the nuclei decay. Nuclear decay is an example of a purely statistical process. A more precise definition of half-life is that *each nucleus has a 50% chance of living for a time equal to one half-life* $t_{1/2}$. Thus, if N is reasonably large, half of the original nuclei decay in a time of one half-life. If an individual nucleus makes it through that time, it still has a 50% chance of surviving through another half-life. Even if it happens to make it through hundreds of half-lives, it still has a 50% chance of surviving through one more. The probability of decay is the same no matter when you start counting. This is like random coin flipping. The chance of heads is 50%, no matter what has happened before.



Radioactive decay reduces the number of radioactive nuclei over time. In one half-life $t_{1/2}$, the number decreases to half of its original value. Half of what remains decay in the next half-life, and half of those in the next, and so on. This is an exponential decay, as seen in the graph of the number of nuclei present as a function of time.

There is a tremendous range in the half-lives of various nuclides, from as short as 10^{-23} s for the most unstable, to more than 10^{16} y for the least unstable, or about 46 orders of magnitude. Nuclides with the shortest half-lives are those for which the nuclear forces are least attractive, an indication of the extent to which the nuclear force can depend on the particular combination of neutrons and protons. The concept of half-life is applicable to other subatomic particles, as will be discussed in Particle Physics. It is also applicable to the decay of excited states in atoms and nuclei. The following equation gives the quantitative relationship between the original

number of nuclei present at time zero (N_0) and the number (N) at a later time t:

Equation:

$$N = N_0 e^{-\lambda t},$$

where $e = 2.71828...$ is the base of the natural logarithm, and λ is the **decay constant** for the nuclide. The shorter the half-life, the larger is the value of λ, and the faster the exponential $e^{-\lambda t}$ decreases with time. The relationship between the decay constant λ and the half-life $t_{1/2}$ is

Equation:

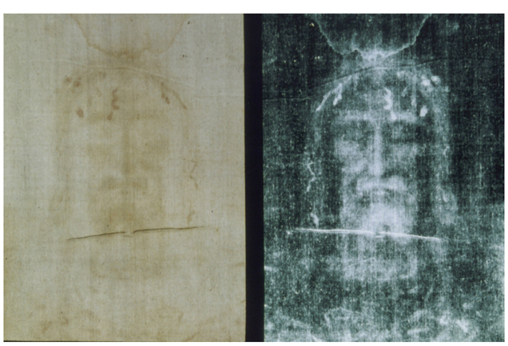
$$\lambda = \frac{\ln(2)}{t_{1/2}} \approx \frac{0.693}{t_{1/2}}.$$

To see how the number of nuclei declines to half its original value in one half-life, let $t = t_{1/2}$ in the exponential in the equation $N = N_0 e^{-\lambda t}$. This gives $N = N_0 e^{-\lambda t} = N_0 e^{-0.693} = 0.500 N_0$. For integral numbers of half-lives, you can just divide the original number by 2 over and over again, rather than using the exponential relationship. For example, if ten half-lives have passed, we divide N by 2 ten times. This reduces it to $N/1024$. For an arbitrary time, not just a multiple of the half-life, the exponential relationship must be used.

Radioactive dating is a clever use of naturally occurring radioactivity. Its most famous application is **carbon-14 dating**. Carbon-14 has a half-life of 5730 years and is produced in a nuclear reaction induced when solar neutrinos strike ^{14}N in the atmosphere. Radioactive carbon has the same chemistry as stable carbon, and so it mixes into the ecosphere, where it is consumed and becomes part of every living organism. Carbon-14 has an abundance of 1.3 parts per trillion of normal carbon. Thus, if you know the number of carbon nuclei in an object (perhaps determined by mass and Avogadro’s number), you multiply that number by 1.3×10^{-12} to find the number of ^{14}C nuclei in the object. When an organism dies, carbon exchange with the environment ceases, and ^{14}C is not replenished as it

decays. By comparing the abundance of ^{14}C in an artifact, such as mummy wrappings, with the normal abundance in living tissue, it is possible to determine the artifact's age (or time since death). Carbon-14 dating can be used for biological tissues as old as 50 or 60 thousand years, but is most accurate for younger samples, since the abundance of ^{14}C nuclei in them is greater. Very old biological materials contain no ^{14}C at all. There are instances in which the date of an artifact can be determined by other means, such as historical knowledge or tree-ring counting. These cross-references have confirmed the validity of carbon-14 dating and permitted us to calibrate the technique as well. Carbon-14 dating revolutionized parts of archaeology and is of such importance that it earned the 1960 Nobel Prize in chemistry for its developer, the American chemist Willard Libby (1908–1980).

One of the most famous cases of carbon-14 dating involves the Shroud of Turin, a long piece of fabric purported to be the burial shroud of Jesus (see [link]). This relic was first displayed in Turin in 1354 and was denounced as a fraud at that time by a French bishop. Its remarkable negative imprint of an apparently crucified body resembles the then-accepted image of Jesus, and so the shroud was never disregarded completely and remained controversial over the centuries. Carbon-14 dating was not performed on the shroud until 1988, when the process had been refined to the point where only a small amount of material needed to be destroyed. Samples were tested at three independent laboratories, each being given four pieces of cloth, with only one unidentified piece from the shroud, to avoid prejudice. All three laboratories found samples of the shroud contain 92% of the ^{14}C found in living tissues, allowing the shroud to be dated (see [link]).



Part of the Shroud of Turin, which shows a remarkable negative imprint likeness of Jesus complete with evidence of crucifixion wounds. The shroud first surfaced in the 14th century and was only recently carbon-14 dated. It has not been determined how the image was placed on the material. (credit: Butko, Wikimedia Commons)

Example:
How Old Is the Shroud of Turin?

Calculate the age of the Shroud of Turin given that the amount of ${}^{14}\text{C}$ found in it is 92% of that in living tissue.

Strategy

Knowing that 92% of the ${}^{14}\text{C}$ remains means that $N/N_0 = 0.92$. Therefore, the equation $N = N_0 e^{-\lambda t}$ can be used to find λt. We also know that the half-life of ${}^{14}\text{C}$ is 5730 y, and so once λt is known, we can use the equation $\lambda = \frac{0.693}{t_{1/2}}$ to find λ and then find t as requested. Here, we postulate that the decrease in ${}^{14}\text{C}$ is solely due to nuclear decay.

Solution

Solving the equation $N = N_0 e^{-\lambda t}$ for N/N_0 gives

Equation:

$$\frac{N}{N_0} = e^{-\lambda t}.$$

Thus,
Equation:

$$0.92 = e^{-\lambda t}.$$

Taking the natural logarithm of both sides of the equation yields
Equation:

$$\ln 0.92 = -\lambda t$$

so that
Equation:

$$-0.0834 = -\lambda t.$$

Rearranging to isolate t gives
Equation:

$$t = \frac{0.0834}{\lambda}.$$

Now, the equation $\lambda = \frac{0.693}{t_{1/2}}$ can be used to find λ for ^{14}C. Solving for λ and substituting the known half-life gives
Equation:

$$\lambda = \frac{0.693}{t_{1/2}} = \frac{0.693}{5730\ \text{y}}.$$

We enter this value into the previous equation to find t:
Equation:

$$t = \frac{0.0834}{\frac{0.693}{5730\ \text{y}}} = 690\ \text{y}.$$

Discussion
This dates the material in the shroud to 1988–690 = a.d. 1300. Our calculation is only accurate to two digits, so that the year is rounded to 1300. The values obtained at the three independent laboratories gave a

weighted average date of a.d. 1320 ± 60. The uncertainty is typical of carbon-14 dating and is due to the small amount of ^{14}C in living tissues, the amount of material available, and experimental uncertainties (reduced by having three independent measurements). It is meaningful that the date of the shroud is consistent with the first record of its existence and inconsistent with the period in which Jesus lived.

There are other forms of radioactive dating. Rocks, for example, can sometimes be dated based on the decay of ^{238}U. The decay series for ^{238}U ends with ^{206}Pb, so that the ratio of these nuclides in a rock is an indication of how long it has been since the rock solidified. The original composition of the rock, such as the absence of lead, must be known with some confidence. However, as with carbon-14 dating, the technique can be verified by a consistent body of knowledge. Since ^{238}U has a half-life of 4.5×10^9 y, it is useful for dating only very old materials, showing, for example, that the oldest rocks on Earth solidified about 3.5×10^9 years ago.

Activity, the Rate of Decay

What do we mean when we say a source is highly radioactive? Generally, this means the number of decays per unit time is very high. We define **activity** R to be the **rate of decay** expressed in decays per unit time. In equation form, this is

Equation:

$$R = \frac{\Delta N}{\Delta t}$$

where ΔN is the number of decays that occur in time Δt. The SI unit for activity is one decay per second and is given the name **becquerel** (Bq) in honor of the discoverer of radioactivity. That is,

Equation:

$$1 \text{ Bq} = 1 \text{ decay/s.}$$

Activity R is often expressed in other units, such as decays per minute or decays per year. One of the most common units for activity is the **curie** (Ci), defined to be the activity of 1 g of ^{226}Ra, in honor of Marie Curie's work with radium. The definition of curie is

Equation:

$$1 \text{ Ci} = 3.70 \times 10^{10} \text{ Bq},$$

or 3.70×10^{10} decays per second. A curie is a large unit of activity, while a becquerel is a relatively small unit. $1 \text{ MBq} = 100 \text{ microcuries } (\mu\text{Ci})$. In countries like Australia and New Zealand that adhere more to SI units, most radioactive sources, such as those used in medical diagnostics or in physics laboratories, are labeled in Bq or megabecquerel (MBq).

Intuitively, you would expect the activity of a source to depend on two things: the amount of the radioactive substance present, and its half-life. The greater the number of radioactive nuclei present in the sample, the more will decay per unit of time. The shorter the half-life, the more decays per unit time, for a given number of nuclei. So activity R should be proportional to the number of radioactive nuclei, N, and inversely proportional to their half-life, $t_{1/2}$. In fact, your intuition is correct. It can be shown that the activity of a source is

Equation:

$$R = \frac{0.693N}{t_{1/2}}$$

where N is the number of radioactive nuclei present, having half-life $t_{1/2}$. This relationship is useful in a variety of calculations, as the next two examples illustrate.

Example:

How Great Is the ^{14}C Activity in Living Tissue?

Calculate the activity due to ^{14}C in 1.00 kg of carbon found in a living organism. Express the activity in units of Bq and Ci.

Strategy

To find the activity R using the equation $R = \frac{0.693N}{t_{1/2}}$, we must know N and $t_{1/2}$. The half-life of ^{14}C can be found in Appendix B, and was stated above as 5730 y. To find N, we first find the number of ^{12}C nuclei in 1.00 kg of carbon using the concept of a mole. As indicated, we then multiply by 1.3×10^{-12} (the abundance of ^{14}C in a carbon sample from a living organism) to get the number of ^{14}C nuclei in a living organism.

Solution

One mole of carbon has a mass of 12.0 g, since it is nearly pure ^{12}C. (A mole has a mass in grams equal in magnitude to A found in the periodic table.) Thus the number of carbon nuclei in a kilogram is

Equation:

$$N(^{12}\text{C}) = \frac{6.02 \times 10^{23}\ \text{mol}^{-1}}{12.0\ \text{g/mol}} \times (1000\ \text{g}) = 5.02 \times 10^{25}.$$

So the number of ^{14}C nuclei in 1 kg of carbon is

Equation:

$$N(^{14}\text{C}) = (5.02 \times 10^{25})(1.3 \times 10^{-12}) = 6.52 \times 10^{13}.$$

Now the activity R is found using the equation $R = \frac{0.693N}{t_{1/2}}$.

Entering known values gives

Equation:

$$R = \frac{0.693(6.52\times10^{13})}{5730\ \text{y}} = 7.89\times10^{9}\ \text{y}^{-1},$$

or 7.89×10^{9} decays per year. To convert this to the unit Bq, we simply convert years to seconds. Thus,

Equation:

$$R = (7.89 \times 10^{9}\ \text{y}^{-1}) \frac{1.00\ \text{y}}{3.16 \times 10^{7}\ \text{s}} = 250\ \text{Bq},$$

or 250 decays per second. To express R in curies, we use the definition of a curie,

Equation:

$$R = \frac{250\ \text{Bq}}{3.7 \times 10^{10}\ \text{Bq/Ci}} = 6.76 \times 10^{-9}\ \text{Ci}.$$

Thus,

Equation:

$$R = 6.76\ \text{nCi}.$$

Discussion

Our own bodies contain kilograms of carbon, and it is intriguing to think there are hundreds of ^{14}C decays per second taking place in us. Carbon-14 and other naturally occurring radioactive substances in our bodies contribute to the background radiation we receive. The small number of decays per second found for a kilogram of carbon in this example gives you some idea of how difficult it is to detect ^{14}C in a small sample of material. If there are 250 decays per second in a kilogram, then there are 0.25 decays per second in a gram of carbon in living tissue. To observe this, you must be able to distinguish decays from other forms of radiation, in order to reduce background noise. This becomes more difficult with an old tissue sample, since it contains less ^{14}C, and for samples more than 50 thousand years old, it is impossible.

Human-made (or artificial) radioactivity has been produced for decades and has many uses. Some of these include medical therapy for cancer, medical imaging and diagnostics, and food preservation by irradiation. Many applications as well as the biological effects of radiation are explored in Medical Applications of Nuclear Physics, but it is clear that radiation is hazardous. A number of tragic examples of this exist, one of the most disastrous being the meltdown and fire at the Chernobyl reactor complex in

the Ukraine (see [link]). Several radioactive isotopes were released in huge quantities, contaminating many thousands of square kilometers and directly affecting hundreds of thousands of people. The most significant releases were of ^{131}I, ^{90}Sr, ^{137}Cs, ^{239}Pu, ^{238}U, and ^{235}U. Estimates are that the total amount of radiation released was about 100 million curies.

Human and Medical Applications

The Chernobyl reactor. More than 100 people died soon after its meltdown, and there will be thousands of deaths from radiation-induced cancer in the future. While the accident was due to a series of human errors, the cleanup efforts were heroic. Most of the immediate fatalities were firefighters and reactor personnel. (credit: Elena Filatova)

Example:

What Mass of ^{137}Cs Escaped Chernobyl?

It is estimated that the Chernobyl disaster released 6.0 MCi of ^{137}Cs into the environment. Calculate the mass of ^{137}Cs released.

Strategy

We can calculate the mass released using Avogadro's number and the concept of a mole if we can first find the number of nuclei N released. Since the activity R is given, and the half-life of ^{137}Cs is found in Appendix B to be 30.2 y, we can use the equation $R = \frac{0.693N}{t_{1/2}}$ to find N.

Solution

Solving the equation $R = \frac{0.693N}{t_{1/2}}$ for N gives

Equation:

$$N = \frac{Rt_{1/2}}{0.693}.$$

Entering the given values yields

Equation:

$$N = \frac{(6.0\text{ MCi})(30.2\text{ y})}{0.693}.$$

Converting curies to becquerels and years to seconds, we get

Equation:

$$\begin{aligned} N &= \frac{(6.0\times10^{6}\text{ Ci})(3.7\times10^{10}\text{ Bq/Ci})(30.2\text{ y})(3.16\times10^{7}\text{ s/y})}{0.693} \\ &= 3.1\times10^{26}. \end{aligned}$$

One mole of a nuclide ^{A}X has a mass of A grams, so that one mole of ^{137}Cs has a mass of 137 g. A mole has 6.02×10^{23} nuclei. Thus the mass of ^{137}Cs released was

Equation:

$$\begin{aligned} m &= \left(\frac{137\text{ g}}{6.02\times10^{23}}\right)(3.1\times10^{26}) = 70\times10^{3}\text{ g} \\ &= 70\text{ kg}. \end{aligned}$$

Discussion

While 70 kg of material may not be a very large mass compared to the amount of fuel in a power plant, it is extremely radioactive, since it only has a 30-year half-life. Six megacuries (6.0 MCi) is an extraordinary amount of activity but is only a fraction of what is produced in nuclear reactors. Similar amounts of the other isotopes were also released at Chernobyl. Although the chances of such a disaster may have seemed small, the consequences were extremely severe, requiring greater caution than was used. More will be said about safe reactor design in the next chapter, but it should be noted that Western reactors have a fundamentally safer design.

Activity R decreases in time, going to half its original value in one half-life, then to one-fourth its original value in the next half-life, and so on. Since $R=\frac{0.693N}{t_{1/2}}$, the activity decreases as the number of radioactive nuclei decreases. The equation for R as a function of time is found by combining the equations $N=N_0e^{-\lambda t}$ and $R=\frac{0.693N}{t_{1/2}}$, yielding

Equation:

$$R=R_0e^{-\lambda t},$$

where R_0 is the activity at $t=0$. This equation shows exponential decay of radioactive nuclei. For example, if a source originally has a 1.00-mCi activity, it declines to 0.500 mCi in one half-life, to 0.250 mCi in two half-lives, to 0.125 mCi in three half-lives, and so on. For times other than whole half-lives, the equation $R=R_0e^{-\lambda t}$ must be used to find R.

Note:

PhET Explorations: Alpha Decay

Watch alpha particles escape from a polonium nucleus, causing radioactive alpha decay. See how random decay times relate to the half life.

Alpha Deca y.

Section Summary

- Half-life $t_{1/2}$ is the time in which there is a 50% chance that a nucleus will decay. The number of nuclei N as a function of time is
  Equation:

  $$N = N_0 e^{-\lambda t},$$

  where N_0 is the number present at $t = 0$, and λ is the decay constant, related to the half-life by
  Equation:

  $$\lambda = \frac{0.693}{t_{1/2}}.$$

- One of the applications of radioactive decay is radioactive dating, in which the age of a material is determined by the amount of radioactive decay that occurs. The rate of decay is called the activity R:
  Equation:

  $$R = \frac{\Delta N}{\Delta t}.$$

- The SI unit for R is the becquerel (Bq), defined by
  Equation:

  $$1 \text{ Bq} = 1 \text{ decay/s}.$$

- R is also expressed in terms of curies (Ci), where

Equation:

$$1\text{ Ci} = 3.70 \times 10^{10}\text{ Bq.}$$

- The activity R of a source is related to N and $t_{1/2}$ by
  Equation:

$$R = \frac{0.693N}{t_{1/2}}.$$

- Since N has an exponential behavior as in the equation $N = N_0 e^{-\lambda t}$, the activity also has an exponential behavior, given by
  Equation:

$$R = R_0 e^{-\lambda t},$$

  where R_0 is the activity at $t = 0$.

Conceptual Questions

Exercise:

Problem:

In a 3×10^9-year-old rock that originally contained some ^{238}U, which has a half-life of 4.5×10^9 years, we expect to find some ^{238}U remaining in it. Why are ^{226}Ra, ^{222}Rn, and ^{210}Po also found in such a rock, even though they have much shorter half-lives (1600 years, 3.8 days, and 138 days, respectively)?

Exercise:

Problem:

Does the number of radioactive nuclei in a sample decrease to *exactly* half its original value in one half-life? Explain in terms of the statistical nature of radioactive decay.

Exercise:

Problem:

Radioactivity depends on the nucleus and not the atom or its chemical state. Why, then, is one kilogram of uranium more radioactive than one kilogram of uranium hexafluoride?

Exercise:

Problem:

Explain how a bound system can have less mass than its components. Why is this not observed classically, say for a building made of bricks?

Exercise:

Problem:

Spontaneous radioactive decay occurs only when the decay products have less mass than the parent, and it tends to produce a daughter that is more stable than the parent. Explain how this is related to the fact that more tightly bound nuclei are more stable. (Consider the binding energy per nucleon.)

Exercise:

Problem:

To obtain the most precise value of BE from the equation $\text{BE}=\left[ZM\left(^{1}\text{H}\right)+Nm_{n}\right]c^{2}-m\left(^{A}\text{X}\right)c^{2}$, we should take into account the binding energy of the electrons in the neutral atoms. Will doing this produce a larger or smaller value for BE? Why is this effect usually negligible?

Exercise:

Problem:

How does the finite range of the nuclear force relate to the fact that BE/A is greatest for A near 60?

Problems & Exercises

Data from the appendices and the periodic table may be needed for these problems.

Exercise:

Problem:

An old campfire is uncovered during an archaeological dig. Its charcoal is found to contain less than 1/1000 the normal amount of ^{14}C. Estimate the minimum age of the charcoal, noting that $2^{10} = 1024$.

Solution:

57,300 y

Exercise:

Problem:

A ^{60}Co source is labeled 4.00 mCi, but its present activity is found to be 1.85×10^{7} Bq. (a) What is the present activity in mCi? (b) How long ago did it actually have a 4.00-mCi activity?

Exercise:

Problem:

(a) Calculate the activity R in curies of 1.00 g of ^{226}Ra. (b) Discuss why your answer is not exactly 1.00 Ci, given that the curie was originally supposed to be exactly the activity of a gram of radium.

Solution:

(a) 0.988 Ci

(b) The half-life of ^{226}Ra is now better known.

Exercise:

Problem:

Show that the activity of the $^{14}\mathrm{C}$ in 1.00 g of $^{12}\mathrm{C}$ found in living tissue is 0.250 Bq.

Exercise:

Problem:

Mantles for gas lanterns contain thorium, because it forms an oxide that can survive being heated to incandescence for long periods of time. Natural thorium is almost 100% $^{232}\mathrm{Th}$, with a half-life of 1.405×10^{10} y. If an average lantern mantle contains 300 mg of thorium, what is its activity?

Solution:

1.22×10^{3} Bq

Exercise:

Problem:

Cow's milk produced near nuclear reactors can be tested for as little as 1.00 pCi of $^{131}\mathrm{I}$ per liter, to check for possible reactor leakage. What mass of $^{131}\mathrm{I}$ has this activity?

Exercise:

Problem:

(a) Natural potassium contains $^{40}\mathrm{K}$, which has a half-life of 1.277×10^{9} y. What mass of $^{40}\mathrm{K}$ in a person would have a decay rate of 4140 Bq? (b) What is the fraction of $^{40}\mathrm{K}$ in natural potassium, given that the person has 140 g in his body? (These numbers are typical for a 70-kg adult.)

Solution:

(a) 16.0 mg

(b) 0.0114%

Exercise:

Problem:

There is more than one isotope of natural uranium. If a researcher isolates 1.00 mg of the relatively scarce ^{235}U and finds this mass to have an activity of 80.0 Bq, what is its half-life in years?

Exercise:

Problem:

^{50}V has one of the longest known radioactive half-lives. In a difficult experiment, a researcher found that the activity of 1.00 kg of ^{50}V is 1.75 Bq. What is the half-life in years?

Solution:

1.48×10^{17} y

Exercise:

Problem:

You can sometimes find deep red crystal vases in antique stores, called uranium glass because their color was produced by doping the glass with uranium. Look up the natural isotopes of uranium and their half-lives, and calculate the activity of such a vase assuming it has 2.00 g of uranium in it. Neglect the activity of any daughter nuclides.

Exercise:

Problem:

A tree falls in a forest. How many years must pass before the ^{14}C activity in 1.00 g of the tree’s carbon drops to 1.00 decay per hour?

Solution:

5.6×10^{4} y

Exercise:

Problem:

What fraction of the ^{40}K that was on Earth when it formed 4.5×10^9 years ago is left today?

Exercise:

Problem:

A 5000-Ci ^{60}Co source used for cancer therapy is considered too weak to be useful when its activity falls to 3500 Ci. How long after its manufacture does this happen?

Solution:

2.71 y

Exercise:

Problem:

Natural uranium is 0.7200% ^{235}U and 99.27% ^{238}U. What were the percentages of ^{235}U and ^{238}U in natural uranium when Earth formed 4.5×10^9 years ago?

Exercise:

Problem:

The β^- particles emitted in the decay of ^{3}H (tritium) interact with matter to create light in a glow-in-the-dark exit sign. At the time of manufacture, such a sign contains 15.0 Ci of ^{3}H. (a) What is the mass of the tritium? (b) What is its activity 5.00 y after manufacture?

Solution:

(a) 1.56 mg

(b) 11.3 Ci

Exercise:

Problem:

World War II aircraft had instruments with glowing radium-painted dials (see [link]). The activity of one such instrument was 1.0×10^5 Bq when new. (a) What mass of ^{226}Ra was present? (b) After some years, the phosphors on the dials deteriorated chemically, but the radium did not escape. What is the activity of this instrument 57.0 years after it was made?

Exercise:

Problem:

(a) The ^{210}Po source used in a physics laboratory is labeled as having an activity of $1.0\ \mu\text{Ci}$ on the date it was prepared. A student measures the radioactivity of this source with a Geiger counter and observes 1500 counts per minute. She notices that the source was prepared 120 days before her lab. What fraction of the decays is she observing with her apparatus? (b) Identify some of the reasons that only a fraction of the α s emitted are observed by the detector.

Solution:

(a) 1.23×10^{-3}

(b) Only part of the emitted radiation goes in the direction of the detector. Only a fraction of that causes a response in the detector. Some of the emitted radiation (mostly α particles) is observed within the source. Some is absorbed within the source, some is absorbed by the detector, and some does not penetrate the detector.

Exercise:

Problem:

Armor-piercing shells with depleted uranium cores are fired by aircraft at tanks. (The high density of the uranium makes them effective.) The uranium is called depleted because it has had its $^{235}\mathrm{U}$ removed for reactor use and is nearly pure $^{238}\mathrm{U}$. Depleted uranium has been erroneously called non-radioactive. To demonstrate that this is wrong: (a) Calculate the activity of 60.0 g of pure $^{238}\mathrm{U}$. (b) Calculate the activity of 60.0 g of natural uranium, neglecting the $^{234}\mathrm{U}$ and all daughter nuclides.

Exercise:

Problem:

The ceramic glaze on a red-orange Fiestaware plate is $\mathrm{U_2O_3}$ and contains 50.0 grams of $^{238}\mathrm{U}$, but very little $^{235}\mathrm{U}$. (a) What is the activity of the plate? (b) Calculate the total energy that will be released by the $^{238}\mathrm{U}$ decay. (c) If energy is worth 12.0 cents per $\mathrm{kW \cdot h}$, what is the monetary value of the energy emitted? (These plates went out of production some 30 years ago, but are still available as collectibles.)

Solution:

(a) 1.68×10^{-5} Ci

(b) 8.65×10^{10} J

(c) $\$2.9 \times 10^{3}$

Exercise:

Problem:

Large amounts of depleted uranium (^{238}U) are available as a by-product of uranium processing for reactor fuel and weapons. Uranium is very dense and makes good counter weights for aircraft. Suppose you have a 4000-kg block of ^{238}U. (a) Find its activity. (b) How many calories per day are generated by thermalization of the decay energy? (c) Do you think you could detect this as heat? Explain.

Exercise:

Problem:

The *Galileo* space probe was launched on its long journey past several planets in 1989, with an ultimate goal of Jupiter. Its power source is 11.0 kg of $^{238}\mathrm{Pu}$, a by-product of nuclear weapons plutonium production. Electrical energy is generated thermoelectrically from the heat produced when the 5.59-MeV α particles emitted in each decay crash to a halt inside the plutonium and its shielding. The half-life of $^{238}\mathrm{Pu}$ is 87.7 years. (a) What was the original activity of the $^{238}\mathrm{Pu}$ in becquerel? (b) What power was emitted in kilowatts? (c) What power was emitted 12.0 y after launch? You may neglect any extra energy from daughter nuclides and any losses from escaping γ rays.

Solution:

(a) $6.97 \times 10^{15}\ \mathrm{Bq}$

(b) 6.24 kW

(c) 5.67 kW

Exercise:

Problem: Construct Your Own Problem

Consider the generation of electricity by a radioactive isotope in a space probe, such as described in [link]. Construct a problem in which you calculate the mass of a radioactive isotope you need in order to

supply power for a long space flight. Among the things to consider are the isotope chosen, its half-life and decay energy, the power needs of the probe and the length of the flight.

Exercise:

Problem: Unreasonable Results

A nuclear physicist finds $1.0\ \mu$g of ^{236}U in a piece of uranium ore and assumes it is primordial since its half-life is 2.3×10^7 y. (a) Calculate the amount of ^{236}Uthat would had to have been on Earth when it formed 4.5×10^9 y ago for $1.0\ \mu$g to be left today. (b) What is unreasonable about this result? (c) What assumption is responsible?

Exercise:

Problem: Unreasonable Results

(a) Repeat [link] but include the 0.0055% natural abundance of ^{234}U with its 2.45×10^5 y half-life. (b) What is unreasonable about this result? (c) What assumption is responsible? (d) Where does the ^{234}U come from if it is not primordial?

Exercise:

Problem: Unreasonable Results

The manufacturer of a smoke alarm decides that the smallest current of α radiation he can detect is $1.00\ \mu\text{A}$. (a) Find the activity in curies of an α emitter that produces a $1.00\ \mu\text{A}$ current of α particles. (b) What is unreasonable about this result? (c) What assumption is responsible?

Solution:

(a) 84.5 Ci

(b) An extremely large activity, many orders of magnitude greater than permitted for home use.

(c) The assumption of $1.00\ \mu\mathrm{A}$ is unreasonably large. Other methods can detect much smaller decay rates.

Glossary

becquerel
: SI unit for rate of decay of a radioactive material

half-life
: the time in which there is a 50% chance that a nucleus will decay

radioactive dating
: an application of radioactive decay in which the age of a material is determined by the amount of radioactivity of a particular type that occurs

decay constant
: quantity that is inversely proportional to the half-life and that is used in equation for number of nuclei as a function of time

carbon-14 dating
: a radioactive dating technique based on the radioactivity of carbon-14

activity
: the rate of decay for radioactive nuclides

rate of decay
: the number of radioactive events per unit time

curie
: the activity of 1g of $^{226}\mathrm{Ra}$, equal to $3.70 \times 10^{10}\ \mathrm{Bq}$

Binding Energy

- Define and discuss binding energy.
- Calculate the binding energy per nucleon of a particle.

The more tightly bound a system is, the stronger the forces that hold it together and the greater the energy required to pull it apart. We can therefore learn about nuclear forces by examining how tightly bound the nuclei are. We define the **binding energy** (BE) of a nucleus to be *the energy required to completely disassemble it into separate protons and neutrons*. We can determine the BE of a nucleus from its rest mass. The two are connected through Einstein's famous relationship $E=(\Delta m)c^2$. A bound system has a *smaller* mass than its separate constituents; the more tightly the nucleons are bound together, the smaller the mass of the nucleus.

Imagine pulling a nuclide apart as illustrated in [link]. Work done to overcome the nuclear forces holding the nucleus together puts energy into the system. By definition, the energy input equals the binding energy BE. The pieces are at rest when separated, and so the energy put into them increases their total rest mass compared with what it was when they were glued together as a nucleus. That mass increase is thus $\Delta m=\mathrm{BE}/c^2$. This difference in mass is known as *mass defect*. It implies that the mass of the nucleus is less than the sum of the masses of its constituent protons and neutrons. A nuclide ${}^{A}\mathrm{X}$ has Z protons and N neutrons, so that the difference in mass is

Equation:

$$\Delta m=(Zm_p+\mathrm{Nm}_n)-m_{\mathrm{tot}}.$$

Thus,

Equation:

$$\mathrm{BE}=(\Delta m)c^2=[(\mathrm{Zm}_p+\mathrm{Nm}_n)-m_{\mathrm{tot}}]c^2,$$

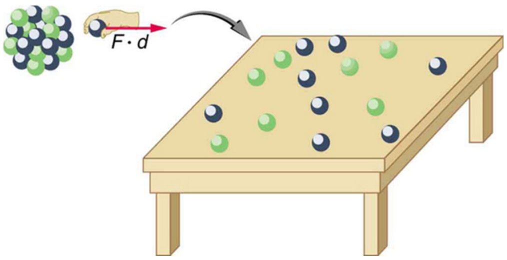
where m_{tot} is the mass of the nuclide ${}^{A}\mathrm{X}$, m_p is the mass of a proton, and m_n is the mass of a neutron. Traditionally, we deal with the masses of

neutral atoms. To get atomic masses into the last equation, we first add Z electrons to m_{tot}, which gives $m\left({}^{A}\text{X}\right)$, the atomic mass of the nuclide. We then add Z electrons to the Z protons, which gives $\text{Zm}\left({}^{1}\text{H}\right)$, or Z times the mass of a hydrogen atom. Thus the binding energy of a nuclide ${}^{A}\text{X}$ is

Equation:

$$\text{BE} = \left\{[Zm({}^{1}\text{H}) + Nm_n] - m({}^{A}X)\right\}c^2.$$

The atomic masses can be found in Appendix A, most conveniently expressed in unified atomic mass units u ($1\ \text{u} = 931.5\ \text{MeV}/c^2$). BE is thus calculated from known atomic masses.



Work done to pull a nucleus apart into its constituent protons and neutrons increases the mass of the system. The work to disassemble the nucleus equals its binding energy BE. A bound system has less mass than the sum of its parts, especially noticeable in the nuclei, where forces and energies are very large.

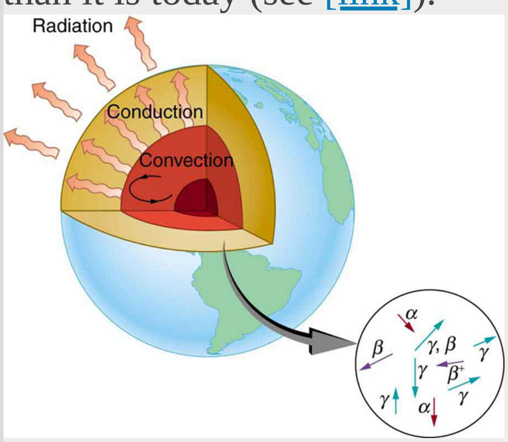
Note:

Things Great and Small

Nuclear Decay Helps Explain Earth’s Hot Interior

A puzzle created by radioactive dating of rocks is resolved by radioactive heating of Earth’s interior. This intriguing story is another example of how small-scale physics can explain large-scale phenomena.

Radioactive dating plays a role in determining the approximate age of the Earth. The oldest rocks on Earth solidified about 3.5×10^9 years ago—a number determined by uranium-238 dating. These rocks could only have solidified once the surface of the Earth had cooled sufficiently. The temperature of the Earth at formation can be estimated based on gravitational potential energy of the assemblage of pieces being converted to thermal energy. Using heat transfer concepts discussed in Thermodynamics it is then possible to calculate how long it would take for the surface to cool to rock-formation temperatures. The result is about 10^9 years. The first rocks formed have been solid for 3.5×10^9 years, so that the age of the Earth is approximately 4.5×10^9 years. There is a large body of other types of evidence (both Earth-bound and solar system characteristics are used) that supports this age. The puzzle is that, given its age and initial temperature, the center of the Earth should be much cooler than it is today (see [link]).



The center of the Earth cools by well-known heat transfer methods. Convection in the liquid regions and conduction

move thermal energy to the surface, where it radiates into cold, dark space. Given the age of the Earth and its initial temperature, it should have cooled to a lower temperature by now. The blowup shows that nuclear decay releases energy in the Earth's interior. This energy has slowed the cooling process and is responsible for the interior still being molten.

We know from seismic waves produced by earthquakes that parts of the interior of the Earth are liquid. Shear or transverse waves cannot travel through a liquid and are not transmitted through the Earth's core. Yet compression or longitudinal waves can pass through a liquid and do go through the core. From this information, the temperature of the interior can be estimated. As noticed, the interior should have cooled more from its initial temperature in the 4.5×10^9 years since its formation. In fact, it should have taken no more than about 10^9 years to cool to its present temperature. What is keeping it hot? The answer seems to be radioactive decay of primordial elements that were part of the material that formed the Earth (see the blowup in [link]).

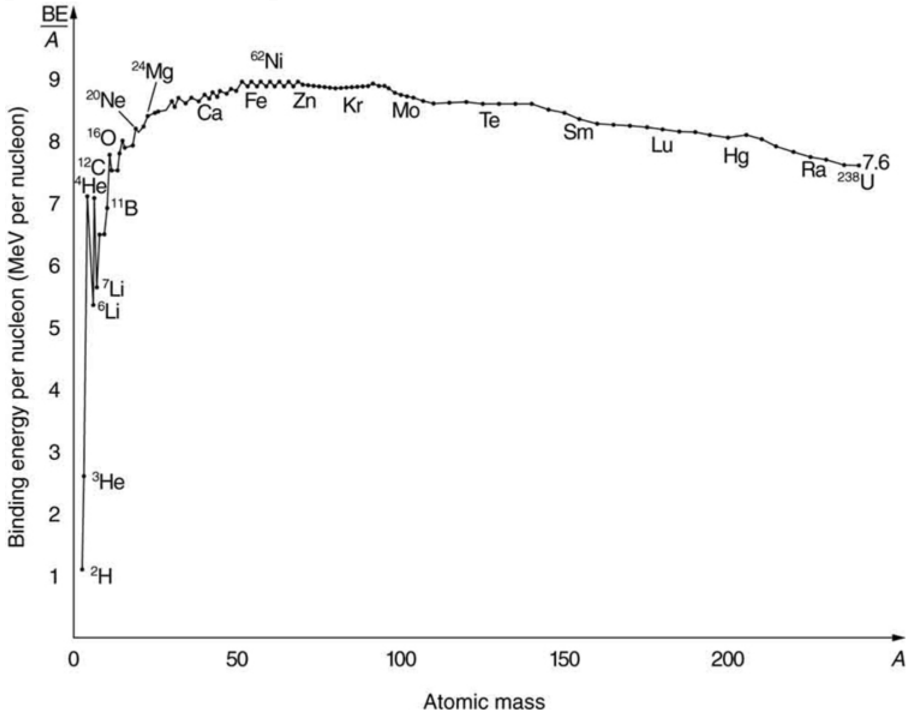
Nuclides such as ^{238}U and ^{40}K have half-lives similar to or longer than the age of the Earth, and their decay still contributes energy to the interior. Some of the primordial radioactive nuclides have unstable decay products that also release energy— ^{238}U has a long decay chain of these. Further, there were more of these primordial radioactive nuclides early in the life of the Earth, and thus the activity and energy contributed were greater then (perhaps by an order of magnitude). The amount of power created by these decays per cubic meter is very small. However, since a huge volume of

material lies deep below the surface, this relatively small amount of energy cannot escape quickly. The power produced near the surface has much less distance to go to escape and has a negligible effect on surface temperatures.

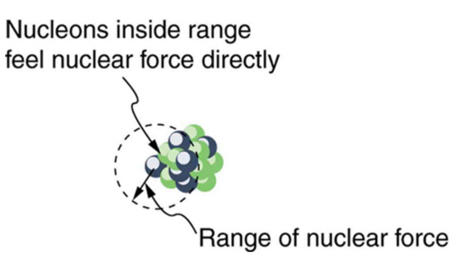
A final effect of this trapped radiation merits mention. Alpha decay produces helium nuclei, which form helium atoms when they are stopped and capture electrons. Most of the helium on Earth is obtained from wells and is produced in this manner. Any helium in the atmosphere will escape in geologically short times because of its high thermal velocity.

What patterns and insights are gained from an examination of the binding energy of various nuclides? First, we find that BE is approximately proportional to the number of nucleons A in any nucleus. About twice as much energy is needed to pull apart a nucleus like $^{24}\mathrm{Mg}$ compared with pulling apart $^{12}\mathrm{C}$, for example. To help us look at other effects, we divide BE by A and consider the **binding energy per nucleon**, BE/A. The graph of BE/A in [link] reveals some very interesting aspects of nuclei. We see that the binding energy per nucleon averages about 8 MeV, but is lower for both the lightest and heaviest nuclei. This overall trend, in which nuclei with A equal to about 60 have the greatest BE/A and are thus the most tightly bound, is due to the combined characteristics of the attractive nuclear forces and the repulsive Coulomb force. It is especially important to note two things—the strong nuclear force is about 100 times stronger than the Coulomb force, *and* the nuclear forces are shorter in range compared to the Coulomb force. So, for low-mass nuclei, the nuclear attraction dominates and each added nucleon forms bonds with all others, causing progressively heavier nuclei to have progressively greater values of BE/A. This continues up to $A \approx 60$, roughly corresponding to the mass number of iron. Beyond that, new nucleons added to a nucleus will be too far from some others to feel their nuclear attraction. Added protons, however, feel the repulsion of all other protons, since the Coulomb force is longer in range. Coulomb repulsion grows for progressively heavier nuclei, but nuclear attraction remains about the same, and so BE/A becomes smaller. This is why stable nuclei heavier than $A \approx 40$ have more neutrons than

protons. Coulomb repulsion is reduced by having more neutrons to keep the protons farther apart (see [link]).



A graph of average binding energy per nucleon, BE/A, for stable nuclei. The most tightly bound nuclei are those with A near 60, where the attractive nuclear force has its greatest effect. At higher A s, the Coulomb repulsion progressively reduces the binding energy per nucleon, because the nuclear force is short ranged. The spikes on the curve are very tightly bound nuclides and indicate shell closures.



The nuclear force is

attractive and stronger than the Coulomb force, but it is short ranged. In low-mass nuclei, each nucleon feels the nuclear attraction of all others. In larger nuclei, the range of the nuclear force, shown for a single nucleon, is smaller than the size of the nucleus, but the Coulomb repulsion from all protons reaches all others. If the nucleus is large enough, the Coulomb repulsion can add to overcome the nuclear attraction.

There are some noticeable spikes on the BE/A graph, which represent particularly tightly bound nuclei. These spikes reveal further details of nuclear forces, such as confirming that closed-shell nuclei (those with magic numbers of protons or neutrons or both) are more tightly bound. The spikes also indicate that some nuclei with even numbers for Z and N, and with $Z = N$, are exceptionally tightly bound. This finding can be correlated with some of the cosmic abundances of the elements. The most common elements in the universe, as determined by observations of atomic spectra from outer space, are hydrogen, followed by $^{4}\mathrm{He}$, with much smaller amounts of $^{12}\mathrm{C}$ and other elements. It should be noted that the heavier elements are created in supernova explosions, while the lighter ones are produced by nuclear fusion during the normal life cycles of stars, as will be discussed in subsequent chapters. The most common elements have the

most tightly bound nuclei. It is also no accident that one of the most tightly bound light nuclei is ^4He, emitted in α decay.

Example:

What Is BE/A for an Alpha Particle?

Calculate the binding energy per nucleon of ^4He, the α particle.

Strategy

To find BE/A, we first find BE using the Equation $\text{BE} = \{[\text{Zm}(^1\text{H}) + \text{Nm}_n] - m(^A\text{X})\}c^2$ and then divide by A. This is straightforward once we have looked up the appropriate atomic masses in Appendix A.

Solution

The binding energy for a nucleus is given by the equation

Equation:

$$\text{BE} = \{[\text{Zm}(^1\text{H}) + \text{Nm}_n] - m(^A\text{X})\}c^2.$$

For ^4He, we have $Z = N = 2$; thus,

Equation:

$$\text{BE} = \{[2m(^1\text{H}) + 2m_n] - m(^4\text{He})\}c^2.$$

Appendix A gives these masses as $m(^4\text{He}) = 4.002602\ \text{u}$, $m(^1\text{H}) = 1.007825\ \text{u}$, and $m_n = 1.008665\ \text{u}$. Thus,

Equation:

$$\text{BE} = (0.030378\ \text{u})c^2.$$

Noting that $1\ \text{u} = 931.5\ \text{MeV}/c^2$, we find

Equation:

$$\text{BE} = (0.030378)(931.5\ \text{MeV}/c^2)c^2 = 28.3\ \text{MeV}.$$

Since $A = 4$, we see that BE/A is this number divided by 4, or

Equation:

$$\text{BE}/A = 7.07\text{ MeV/nucleon.}$$

Discussion
This is a large binding energy per nucleon compared with those for other low-mass nuclei, which have $\text{BE}/A \approx 3\text{ MeV/nucleon}$. This indicates that ^4He is tightly bound compared with its neighbors on the chart of the nuclides. You can see the spike representing this value of BE/A for ^4He on the graph in [link]. This is why ^4He is stable. Since ^4He is tightly bound, it has less mass than other $A = 4$ nuclei and, therefore, cannot spontaneously decay into them. The large binding energy also helps to explain why some nuclei undergo α decay. Smaller mass in the decay products can mean energy release, and such decays can be spontaneous. Further, it can happen that two protons and two neutrons in a nucleus can randomly find themselves together, experience the exceptionally large nuclear force that binds this combination, and act as a ^4He unit within the nucleus, at least for a while. In some cases, the ^4He escapes, and α decay has then taken place.

There is more to be learned from nuclear binding energies. The general trend in BE/A is fundamental to energy production in stars, and to fusion and fission energy sources on Earth, for example. This is one of the applications of nuclear physics covered in Medical Applications of Nuclear Physics. The abundance of elements on Earth, in stars, and in the universe as a whole is related to the binding energy of nuclei and has implications for the continued expansion of the universe.

Problem-Solving Strategies

For Reaction And Binding Energies and Activity Calculations in Nuclear Physics

1. *Identify exactly what needs to be determined in the problem (identify the unknowns).* This will allow you to decide whether the energy of a decay or nuclear reaction is involved, for example, or whether the problem is primarily concerned with activity (rate of decay).

2. *Make a list of what is given or can be inferred from the problem as stated (identify the knowns).*
3. *For reaction and binding-energy problems, we use atomic rather than nuclear masses.* Since the masses of neutral atoms are used, you must count the number of electrons involved. If these do not balance (such as in β^+ decay), then an energy adjustment of 0.511 MeV per electron must be made. Also note that atomic masses may not be given in a problem; they can be found in tables.
4. *For problems involving activity, the relationship of activity to half-life, and the number of nuclei given in the equation* $R = \frac{0.693N}{t_{1/2}}$ *can be very useful.* Owing to the fact that number of nuclei is involved, you will also need to be familiar with moles and Avogadro's number.
5. *Perform the desired calculation; keep careful track of plus and minus signs as well as powers of 10.*
6. *Check the answer to see if it is reasonable: Does it make sense?* Compare your results with worked examples and other information in the text. (Heeding the advice in Step 5 will also help you to be certain of your result.) You must understand the problem conceptually to be able to determine whether the numerical result is reasonable.

Note:
PhET Explorations: Nuclear Fission
Start a chain reaction, or introduce non-radioactive isotopes to prevent one. Control energy production in a nuclear reactor!

https://archive.cnx.org/specials/01caf0d0-116f-11e6-b891-abfdaa77b03b/nuclear-fission/#sim-one-nucleus

Section Summary

- The binding energy (BE) of a nucleus is the energy needed to separate it into individual protons and neutrons. In terms of atomic masses,
  Equation:

$$\mathrm{BE} = \{[\mathrm{Zm}(^1\mathrm{H}) + \mathrm{Nm}_n] - m(^A\mathrm{X})\}c^2,$$

where $m\left(^1\mathrm{H}\right)$ is the mass of a hydrogen atom, $m\left(^A\mathrm{X}\right)$ is the atomic mass of the nuclide, and m_n is the mass of a neutron. Patterns in the binding energy per nucleon, BE/A, reveal details of the nuclear force. The larger the BE/A, the more stable the nucleus.

Conceptual Questions

Exercise:

Problem:

Why is the number of neutrons greater than the number of protons in stable nuclei having A greater than about 40, and why is this effect more pronounced for the heaviest nuclei?

Problems & Exercises

Exercise:

Problem:

$^2\mathrm{H}$ is a loosely bound isotope of hydrogen. Called deuterium or heavy hydrogen, it is stable but relatively rare—it is 0.015% of natural hydrogen. Note that deuterium has $Z = N$, which should tend to make it more tightly bound, but both are odd numbers. Calculate BE/A, the binding energy per nucleon, for $^2\mathrm{H}$ and compare it with the approximate value obtained from the graph in [link].

Solution:

1.112 MeV, consistent with graph

Exercise:

Problem:

$^{56}\mathrm{Fe}$ is among the most tightly bound of all nuclides. It is more than 90% of natural iron. Note that $^{56}\mathrm{Fe}$ has even numbers of both protons and neutrons. Calculate BE/A, the binding energy per nucleon, for $^{56}\mathrm{Fe}$ and compare it with the approximate value obtained from the graph in [link].

Exercise:

Problem:

$^{209}\mathrm{Bi}$ is the heaviest stable nuclide, and its BE/A is low compared with medium-mass nuclides. Calculate BE/A, the binding energy per nucleon, for $^{209}\mathrm{Bi}$ and compare it with the approximate value obtained from the graph in [link].

Solution:

7.848 MeV, consistent with graph

Exercise:

Problem:

(a) Calculate BE/A for $^{235}\mathrm{U}$, the rarer of the two most common uranium isotopes. (b) Calculate BE/A for $^{238}\mathrm{U}$. (Most of uranium is $^{238}\mathrm{U}$.) Note that $^{238}\mathrm{U}$ has even numbers of both protons and neutrons. Is the BE/A of $^{238}\mathrm{U}$ significantly different from that of $^{235}\mathrm{U}$?

Exercise:

Problem:

(a) Calculate BE/A for $^{12}\mathrm{C}$. Stable and relatively tightly bound, this nuclide is most of natural carbon. (b) Calculate BE/A for $^{14}\mathrm{C}$. Is the difference in BE/A between $^{12}\mathrm{C}$ and $^{14}\mathrm{C}$ significant? One is stable and common, and the other is unstable and rare.

Solution:

(a) 7.680 MeV, consistent with graph

(b) 7.520 MeV, consistent with graph. Not significantly different from value for ^{12}C, but sufficiently lower to allow decay into another nuclide that is more tightly bound.

Exercise:

Problem:

The fact that BE/A is greatest for A near 60 implies that the range of the nuclear force is about the diameter of such nuclides. (a) Calculate the diameter of an $A = 60$ nucleus. (b) Compare BE/A for ^{58}Ni and ^{90}Sr. The first is one of the most tightly bound nuclides, while the second is larger and less tightly bound.

Exercise:

Problem:

The purpose of this problem is to show in three ways that the binding energy of the electron in a hydrogen atom is negligible compared with the masses of the proton and electron. (a) Calculate the mass equivalent in u of the 13.6-eV binding energy of an electron in a hydrogen atom, and compare this with the mass of the hydrogen atom obtained from Appendix A. (b) Subtract the mass of the proton given in [link] from the mass of the hydrogen atom given in Appendix A. You will find the difference is equal to the electron's mass to three digits, implying the binding energy is small in comparison. (c) Take the ratio of the binding energy of the electron (13.6 eV) to the energy equivalent of the electron's mass (0.511 MeV). (d) Discuss how your answers confirm the stated purpose of this problem.

Solution:

(a) 1.46×10^{-8} u vs. 1.007825 u for ^{1}H

(b) 0.000549 u

(c) 2.66×10^{-5}

Exercise:

Problem: **Unreasonable Results**

A particle physicist discovers a neutral particle with a mass of 2.02733 u that he assumes is two neutrons bound together. (a) Find the binding energy. (b) What is unreasonable about this result? (c) What assumptions are unreasonable or inconsistent?

Solution:

(a) -9.315 MeV

(b) The negative binding energy implies an unbound system.

(c) This assumption that it is two bound neutrons is incorrect.

Glossary

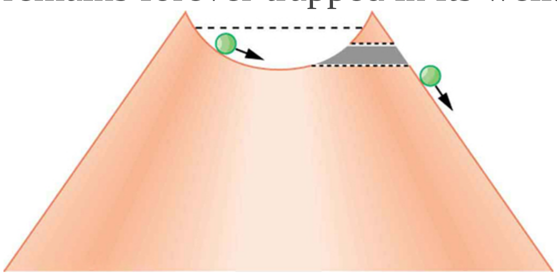
binding energy
: the energy needed to separate nucleus into individual protons and neutrons

binding energy per nucleon
: the binding energy calculated per nucleon; it reveals the details of the nuclear force—larger the BE/A, the more stable the nucleus

Tunneling

- Define and discuss tunneling.
- Define potential barrier.
- Explain quantum tunneling.

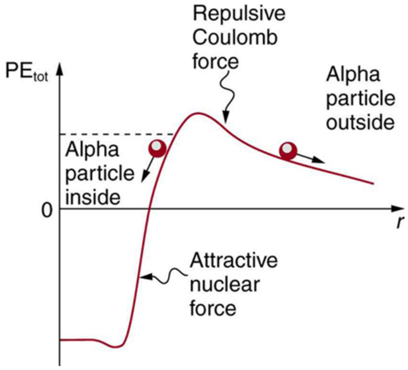
Protons and neutrons are *bound* inside nuclei, that means energy must be supplied to break them away. The situation is analogous to a marble in a bowl that can roll around but lacks the energy to get over the rim. It is bound inside the bowl (see [link]). If the marble could get over the rim, it would gain kinetic energy by rolling down outside. However classically, if the marble does not have enough kinetic energy to get over the rim, it remains forever trapped in its well.



The marble in this semicircular bowl at the top of a volcano has enough kinetic energy to get to the altitude of the dashed line, but not enough to get over the rim, so that it is trapped forever. If it could find a tunnel through the barrier, it would escape, roll downhill, and gain kinetic energy.

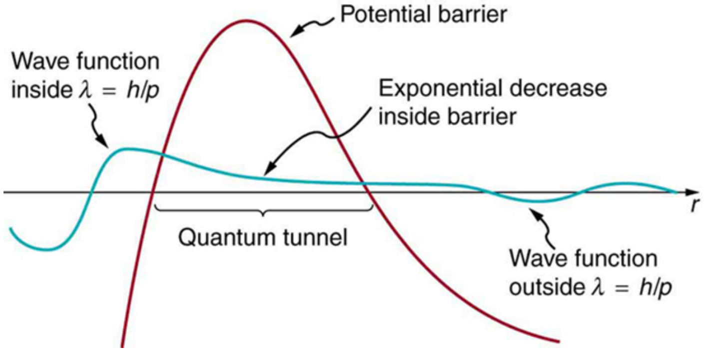
In a nucleus, the attractive nuclear potential is analogous to the bowl at the top of a volcano (where the "volcano" refers only to the shape). Protons and neutrons have kinetic energy, but it is about 8 MeV less than that needed to get out (see [link]). That is, they are bound by an average of 8 MeV per

nucleon. The slope of the hill outside the bowl is analogous to the repulsive Coulomb potential for a nucleus, such as for an α particle outside a positive nucleus. In α decay, two protons and two neutrons spontaneously break away as a $^4\mathrm{He}$ unit. Yet the protons and neutrons do not have enough kinetic energy to get over the rim. So how does the α particle get out?



Nucleons within an atomic nucleus are bound or trapped by the attractive nuclear force, as shown in this simplified potential energy curve. An α particle outside the range of the nuclear force feels the repulsive Coulomb force. The α particle inside the nucleus does not have enough kinetic energy to get over the rim, yet it does manage to get out by quantum mechanical tunneling.

The answer was supplied in 1928 by the Russian physicist George Gamow (1904–1968). The α particle tunnels through a region of space it is forbidden to be in, and it comes out of the side of the nucleus. Like an electron making a transition between orbits around an atom, it travels from one point to another without ever having been in between. [link] indicates how this works. The wave function of a quantum mechanical particle varies smoothly, going from within an atomic nucleus (on one side of a potential energy barrier) to outside the nucleus (on the other side of the potential energy barrier). Inside the barrier, the wave function does not become zero but decreases exponentially, and we do not observe the particle inside the barrier. The probability of finding a particle is related to the square of its wave function, and so there is a small probability of finding the particle outside the barrier, which implies that the particle can tunnel through the barrier. This process is called **barrier penetration** or **quantum mechanical tunneling**. This concept was developed in theory by J. Robert Oppenheimer (who led the development of the first nuclear bombs during World War II) and was used by Gamow and others to describe α decay.

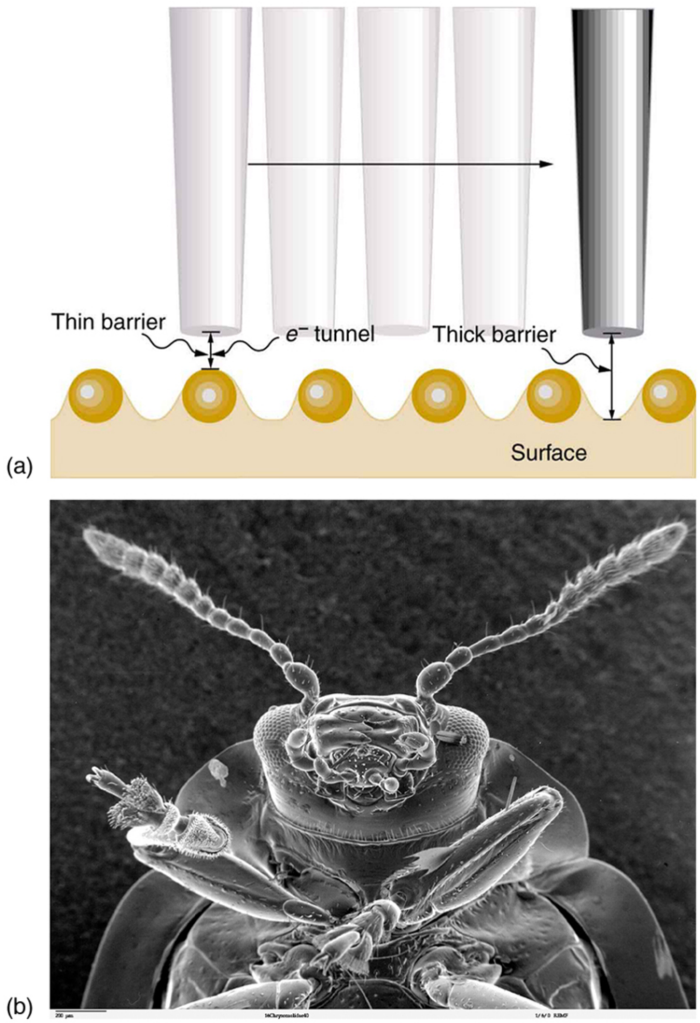


The wave function representing a quantum mechanical particle must vary smoothly, going from within the nucleus (to the left of the barrier) to outside the nucleus (to the right of the barrier). Inside the barrier, the wave function does not abruptly become zero; rather, it decreases exponentially. Outside the barrier, the

wave function is small but finite, and there it smoothly becomes sinusoidal. Owing to the fact that there is a small probability of finding the particle outside the barrier, the particle can tunnel through the barrier.

Good ideas explain more than one thing. In addition to qualitatively explaining how the four nucleons in an α particle can get out of the nucleus, the detailed theory also explains quantitatively the half-life of various nuclei that undergo α decay. This description is what Gamow and others devised, and it works for α decay half-lives that vary by 17 orders of magnitude. Experiments have shown that the more energetic the α decay of a particular nuclide is, the shorter is its half-life. **Tunneling** explains this in the following manner: For the decay to be more energetic, the nucleons must have more energy in the nucleus and should be able to ascend a little closer to the rim. The barrier is therefore not as thick for more energetic decay, and the exponential decrease of the wave function inside the barrier is not as great. Thus the probability of finding the particle outside the barrier is greater, and the half-life is shorter.

Tunneling as an effect also occurs in quantum mechanical systems other than nuclei. Electrons trapped in solids can tunnel from one object to another if the barrier between the objects is thin enough. The process is the same in principle as described for α decay. It is far more likely for a thin barrier than a thick one. Scanning tunneling electron microscopes function on this principle. The current of electrons that travels between a probe and a sample tunnels through a barrier and is very sensitive to its thickness, allowing detection of individual atoms as shown in [link].



(a) A scanning tunneling electron microscope can detect extremely small variations in dimensions, such as individual atoms. Electrons tunnel quantum mechanically between the probe and the sample. The probability of tunneling is extremely sensitive to barrier thickness, so that the electron current is a sensitive indicator of surface features. (b) Head and mouthparts of *Coleoptera Chrysomelidea* as seen through an electron microscope (credit: Louisa Howard, Dartmouth College)

Note:
PhET Explorations: Quantum Tunneling and Wave Packets
Watch quantum "particles" tunnel through barriers. Explore the properties of the wave functions that describe these particles.

Quantum Tunneling and Wave Packets

Section Summary

- Tunneling is a quantum mechanical process of potential energy barrier penetration. The concept was first applied to explain α decay, but tunneling is found to occur in other quantum mechanical systems.

Conceptual Questions

Exercise:

Problem:

A physics student caught breaking conservation laws is imprisoned. She leans against the cell wall hoping to tunnel out quantum mechanically. Explain why her chances are negligible. (This is so in any classical situation.)

Exercise:

Problem:

When a nucleus α decays, does the α particle move continuously from inside the nucleus to outside? That is, does it travel each point along an imaginary line from inside to out? Explain.

Problems-Exercises

Exercise:

Problem:

Derive an approximate relationship between the energy of α decay and half-life using the following data. It may be useful to graph the log of $t_{1/2}$ against E_α to find some straight-line relationship.

Nuclide	E_α (MeV)	$t_{1/2}$
^{216}Ra	9.5	0.18 μs
^{194}Po	7.0	0.7 s
^{240}Cm	6.4	27 d
^{226}Ra	4.91	1600 y
^{232}Th	4.1	1.4×10^{10} y

Energy and Half-Life for α Decay

Exercise:

Problem: Integrated Concepts

A 2.00-T magnetic field is applied perpendicular to the path of charged particles in a bubble chamber. What is the radius of curvature of the path of a 10 MeV proton in this field? Neglect any slowing along its path.

Solution:

22.8 cm

Exercise:

Problem:

(a) Write the decay equation for the α decay of ^{235}U. (b) What energy is released in this decay? The mass of the daughter nuclide is 231.036298 u. (c) Assuming the residual nucleus is formed in its ground state, how much energy goes to the α particle?

Solution:

(a) ${}^{235}_{92}\text{U}_{143} \rightarrow {}^{231}_{90}\text{Th}_{141} + {}^{4}_{2}\text{He}_{2}$

(b) 4.679 MeV

(c) 4.599 MeV

Exercise:

Problem: Unreasonable Results

The relatively scarce naturally occurring calcium isotope ^{48}Ca has a half-life of about 2×10^{16} y. (a) A small sample of this isotope is labeled as having an activity of 1.0 Ci. What is the mass of the ^{48}Ca in the sample? (b) What is unreasonable about this result? (c) What assumption is responsible?

Exercise:

Problem: Unreasonable Results

A physicist scatters γ rays from a substance and sees evidence of a nucleus 7.5×10^{-13} m in radius. (a) Find the atomic mass of such a nucleus. (b) What is unreasonable about this result? (c) What is unreasonable about the assumption?

Solution:

a) 2.4×10^{8} u

(b) The greatest known atomic masses are about 260. This result found in (a) is extremely large.

(c) The assumed radius is much too large to be reasonable.

Exercise:

Problem: Unreasonable Results

A frazzled theoretical physicist reckons that all conservation laws are obeyed in the decay of a proton into a neutron, positron, and neutrino (as in β^{+} decay of a nucleus) and sends a paper to a journal to announce the reaction as a possible end of the universe due to the spontaneous decay of protons. (a) What energy is released in this decay? (b) What is unreasonable about this result? (c) What assumption is responsible?

Solution:

(a) –1.805 MeV

(b) Negative energy implies energy input is necessary and the reaction cannot be spontaneous.

(c) Although all conversation laws are obeyed, energy must be supplied, so the assumption of spontaneous decay is incorrect.

Exercise:

Problem: Construct Your Own Problem

Consider the decay of radioactive substances in the Earth's interior. The energy emitted is converted to thermal energy that reaches the earth's surface and is radiated away into cold dark space. Construct a problem in which you estimate the activity in a cubic meter of earth rock? And then calculate the power generated. Calculate how much power must cross each square meter of the Earth's surface if the power is dissipated at the same rate as it is generated. Among the things to consider are the activity per cubic meter, the energy per decay, and the size of the Earth.

Glossary

barrier penetration
: quantum mechanical effect whereby a particle has a nonzero probability to cross through a potential energy barrier despite not having sufficient energy to pass over the barrier; also called quantum mechanical tunneling

quantum mechanical tunneling
: quantum mechanical effect whereby a particle has a nonzero probability to cross through a potential energy barrier despite not having sufficient energy to pass over the barrier; also called barrier penetration

tunneling
: a quantum mechanical process of potential energy barrier penetration

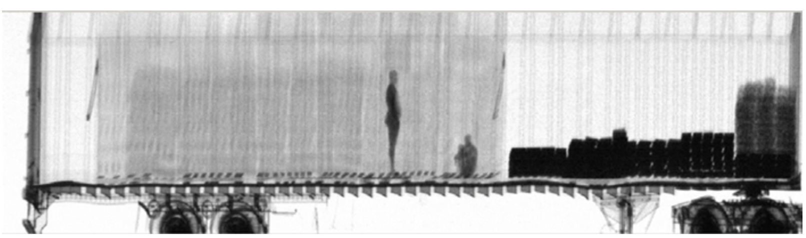
Introduction to Applications of Nuclear Physics
class="introduction"

- Provide examples of various nuclear physics applications.

Tori Randall, Ph.D., curator for the Department of Physical Anthropology at the San Diego Museum of Man, prepares a 550-year-old Peruvian child mummy for a CT scan at Naval Medical Center San Diego. (credit: U.S. Navy photo by Mass Communication Specialist 3rd Class Samantha A. Lewis)

Applications of nuclear physics have become an integral part of modern life. From the bone scan that detects a cancer to the radioiodine treatment that cures another, nuclear radiation has diagnostic and therapeutic effects on medicine. From the fission power reactor to the hope of controlled fusion, nuclear energy is now commonplace and is a part of our plans for the future. Yet, the destructive potential of nuclear weapons haunts us, as does the possibility of nuclear reactor accidents. Certainly, several applications of nuclear physics escape our view, as seen in [link]. Not only has nuclear physics revealed secrets of nature, it has an inevitable impact based on its applications, as they are intertwined with human values. Because of its potential for alleviation of suffering, and its power as an ultimate destructor of life, nuclear physics is often viewed with ambivalence. But it provides perhaps the best example that applications can be good or evil, while knowledge itself is neither.

Customs officers inspect vehicles using neutron irradiation. Cars and trucks pass through portable x-ray machines that reveal their contents. (credit: Gerald L. Nino, CBP, U.S. Dept. of Homeland Security)

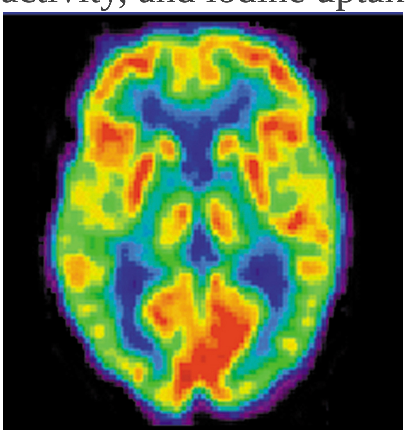


This image shows two stowaways caught illegally entering the United States from Canada. (credit: U.S. Customs and Border Protection)

Medical Imaging and Diagnostics

- Explain the working principle behind an anger camera.
- Describe the SPECT and PET imaging techniques.

A host of medical imaging techniques employ nuclear radiation. What makes nuclear radiation so useful? First, γ radiation can easily penetrate tissue; hence, it is a useful probe to monitor conditions inside the body. Second, nuclear radiation depends on the nuclide and not on the chemical compound it is in, so that a radioactive nuclide can be put into a compound designed for specific purposes. The compound is said to be **tagged**. A tagged compound used for medical purposes is called a **radiopharmaceutical**. Radiation detectors external to the body can determine the location and concentration of a radiopharmaceutical to yield medically useful information. For example, certain drugs are concentrated in inflamed regions of the body, and this information can aid diagnosis and treatment as seen in [link]. Another application utilizes a radiopharmaceutical which the body sends to bone cells, particularly those that are most active, to detect cancerous tumors or healing points. Images can then be produced of such bone scans. Radioisotopes are also used to determine the functioning of body organs, such as blood flow, heart muscle activity, and iodine uptake in the thyroid gland.



A radiopharmaceutical is used to produce this brain image of a patient with

Alzheimer’s disease. Certain features are computer enhanced. (credit: National Institutes of Health)

Medical Application

[link] lists certain medical diagnostic uses of radiopharmaceuticals, including isotopes and activities that are typically administered. Many organs can be imaged with a variety of nuclear isotopes replacing a stable element by a radioactive isotope. One common diagnostic employs iodine to image the thyroid, since iodine is concentrated in that organ. The most active thyroid cells, including cancerous cells, concentrate the most iodine and, therefore, emit the most radiation. Conversely, hypothyroidism is indicated by lack of iodine uptake. Note that there is more than one isotope that can be used for several types of scans. Another common nuclear diagnostic is the thallium scan for the cardiovascular system, particularly used to evaluate blockages in the coronary arteries and examine heart activity. The salt TlCl can be used, because it acts like NaCl and follows the blood. Gallium-67 accumulates where there is rapid cell growth, such as in tumors and sites of infection. Hence, it is useful in cancer imaging. Usually, the patient receives the injection one day and has a whole body scan 3 or 4 days later because it can take several days for the gallium to build up.

Procedure, isotope	**Typical activity (mCi), where** $1\ \text{mCi} = 3.7 \times 10^7\ \text{Bq}$

Procedure, isotope	Typical activity (mCi), where $1 \text{ mCi} = 3.7 \times 10^7 \text{ Bq}$
Brain scan	
^{99m}Tc	7.5
^{113m}In	7.5
^{11}C (PET)	20
^{13}N (PET)	20
^{15}O (PET)	50
^{18}F (PET)	10
Lung scan	
^{99m}Tc	2

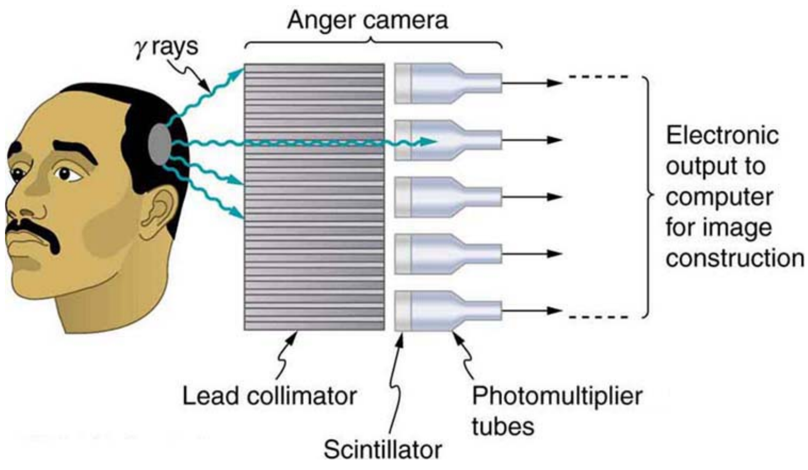
Procedure, isotope	Typical activity (mCi), where $1\ \mathrm{mCi} = 3.7 \times 10^7\ \mathrm{Bq}$
$^{133}\mathrm{Xe}$	7.5
Cardiovascular blood pool	
$^{131}\mathrm{I}$	0.2
$^{99\mathrm{m}}\mathrm{Tc}$	2
Cardiovascular arterial flow	
$^{201}\mathrm{Tl}$	3
$^{24}\mathrm{Na}$	7.5
Thyroid scan	
$^{131}\mathrm{I}$	0.05
$^{123}\mathrm{I}$	0.07

Procedure, isotope	**Typical activity (mCi), where** $1\ \mathrm{mCi} = 3.7 \times 10^7\ \mathrm{Bq}$
Liver scan	
$^{198}\mathrm{Au}$ (colloid)	0.1
$^{99\mathrm{m}}\mathrm{Tc}$ (colloid)	2
Bone scan	
$^{85}\mathrm{Sr}$	0.1
$^{99\mathrm{m}}\mathrm{Tc}$	10
Kidney scan	
$^{197}\mathrm{Hg}$	0.1
$^{99\mathrm{m}}\mathrm{Tc}$	1.5

Diagnostic Uses of Radiopharmaceuticals

Note that [link] lists many diagnostic uses for ^{99m}Tc, where “m” stands for a metastable state of the technetium nucleus. Perhaps 80 percent of all radiopharmaceutical procedures employ ^{99m}Tc because of its many advantages. One is that the decay of its metastable state produces a single, easily identified 0.142-MeV γ ray. Additionally, the radiation dose to the patient is limited by the short 6.0-h half-life of ^{99m}Tc. And, although its half-life is short, it is easily and continuously produced on site. The basic process for production is neutron activation of molybdenum, which quickly β decays into ^{99m}Tc. Technetium-99m can be attached to many compounds to allow the imaging of the skeleton, heart, lungs, kidneys, etc.

[link] shows one of the simpler methods of imaging the concentration of nuclear activity, employing a device called an **Anger camera** or **gamma camera**. A piece of lead with holes bored through it collimates γ rays emerging from the patient, allowing detectors to receive γ rays from specific directions only. The computer analysis of detector signals produces an image. One of the disadvantages of this detection method is that there is no depth information (i.e., it provides a two-dimensional view of the tumor as opposed to a three-dimensional view), because radiation from any location under that detector produces a signal.



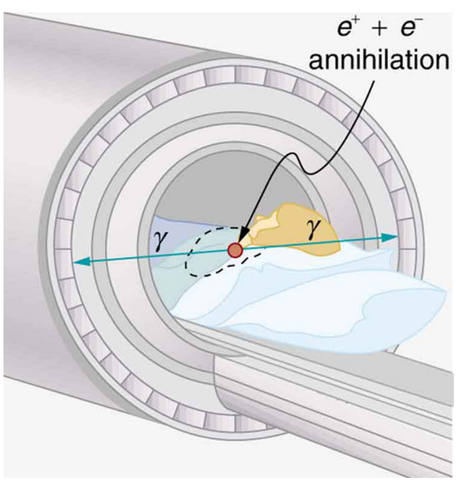
An Anger or gamma camera consists of a lead collimator and an array of detectors. Gamma rays produce light flashes in the

scintillators. The light output is converted to an electrical signal by the photomultipliers. A computer constructs an image from the detector output.

Imaging techniques much like those in x-ray computed tomography (CT) scans use nuclear activity in patients to form three-dimensional images. [link] shows a patient in a circular array of detectors that may be stationary or rotated, with detector output used by a computer to construct a detailed image. This technique is called **single-photon-emission computed tomography(SPECT)** or sometimes simply SPET. The spatial resolution of this technique is poor, about 1 cm, but the contrast (i.e. the difference in visual properties that makes an object distinguishable from other objects and the background) is good.

SPECT uses a geometry similar to a CT scanner to form an image of the concentration of a radiopharmaceutical compound. (credit: Woldo, Wikimedia Commons)

Images produced by β^+ emitters have become important in recent years. When the emitted positron (β^+) encounters an electron, mutual annihilation occurs, producing two γ rays. These γ rays have identical 0.511-MeV energies (the energy comes from the destruction of an electron or positron mass) and they move directly away from one another, allowing detectors to determine their point of origin accurately, as shown in [link]. The system is called **positron emission tomography (PET)**. It requires detectors on opposite sides to simultaneously (i.e., at the same time) detect photons of 0.511-MeV energy and utilizes computer imaging techniques similar to those in SPECT and CT scans. Examples of β^+ -emitting isotopes used in PET are ^{11}C, ^{13}N, ^{15}O, and ^{18}F, as seen in [link]. This list includes C, N, and O, and so they have the advantage of being able to function as tags for natural body compounds. Its resolution of 0.5 cm is better than that of SPECT; the accuracy and sensitivity of PET scans make them useful for examining the brain's anatomy and function. The brain's use of oxygen and water can be monitored with ^{15}O. PET is used extensively for diagnosing brain disorders. It can note decreased metabolism in certain regions prior to a confirmation of Alzheimer's disease. PET can locate regions in the brain that become active when a person carries out specific activities, such as speaking, closing their eyes, and so on.



A PET system takes

advantage of the two identical γ-ray photons produced by positron-electron annihilation. These γ rays are emitted in opposite directions, so that the line along which each pair is emitted is determined. Various events detected by several pairs of detectors are then analyzed by the computer to form an accurate image.

Note:
PhET Explorations: Simplified MRI
Is it a tumor? Magnetic Resonance Imaging (MRI) can tell. Your head is full of tiny radio transmitters (the nuclear spins of the hydrogen nuclei of your water molecules). In an MRI unit, these little radios can be made to broadcast their positions, giving a detailed picture of the inside of your head.

Simplified MRI

Section Summary

- Radiopharmaceuticals are compounds that are used for medical imaging and therapeutics.
- The process of attaching a radioactive substance is called tagging.
- [link] lists certain diagnostic uses of radiopharmaceuticals including the isotope and activity typically used in diagnostics.
- One common imaging device is the Anger camera, which consists of a lead collimator, radiation detectors, and an analysis computer.
- Tomography performed with γ-emitting radiopharmaceuticals is called SPECT and has the advantages of x-ray CT scans coupled with organ- and function-specific drugs.
- PET is a similar technique that uses β^+ emitters and detects the two annihilation γ rays, which aid to localize the source.

Conceptual Questions

Exercise:

Problem:

In terms of radiation dose, what is the major difference between medical diagnostic uses of radiation and medical therapeutic uses?

Exercise:

Problem:

One of the methods used to limit radiation dose to the patient in medical imaging is to employ isotopes with short half-lives. How would this limit the dose?

Problems & Exercises

Exercise:

Problem:

A neutron generator uses an α source, such as radium, to bombard beryllium, inducing the reaction ${}^{4}\text{He} + {}^{9}\text{Be} \rightarrow {}^{12}\text{C} + n$. Such neutron sources are called RaBe sources, or PuBe sources if they use plutonium to get the α s. Calculate the energy output of the reaction in MeV.

Solution:

5.701 MeV

Exercise:

Problem:

Neutrons from a source (perhaps the one discussed in the preceding problem) bombard natural molybdenum, which is 24 percent ${}^{98}\text{Mo}$. What is the energy output of the reaction ${}^{98}\text{Mo} + n \rightarrow {}^{99}\text{Mo} + \gamma$? The mass of ${}^{98}\text{Mo}$ is given in Appendix A: Atomic Masses, and that of ${}^{99}\text{Mo}$ is 98.907711 u.

Exercise:

Problem:

The purpose of producing ${}^{99}\text{Mo}$ (usually by neutron activation of natural molybdenum, as in the preceding problem) is to produce ${}^{99\text{m}}\text{Tc}$. Using the rules, verify that the β^{-} decay of ${}^{99}\text{Mo}$ produces ${}^{99\text{m}}\text{Tc}$. (Most ${}^{99\text{m}}\text{Tc}$ nuclei produced in this decay are left in a metastable excited state denoted ${}^{99\text{m}}\text{Tc}$.)

Solution:

${}^{99}_{42}\text{Mo}_{57} \rightarrow {}^{99}_{43}\text{Tc}_{56} + \beta^{-} + v_e$

Exercise:

Problem:

(a) Two annihilation γ rays in a PET scan originate at the same point and travel to detectors on either side of the patient. If the point of origin is 9.00 cm closer to one of the detectors, what is the difference in arrival times of the photons? (This could be used to give position information, but the time difference is small enough to make it difficult.)

(b) How accurately would you need to be able to measure arrival time differences to get a position resolution of 1.00 mm?

Exercise:

Problem:

[link] indicates that 7.50 mCi of ^{99m}Tc is used in a brain scan. What is the mass of technetium?

Solution:

1.43×10^{-9} g

Exercise:

Problem:

The activities of ^{131}I and ^{123}I used in thyroid scans are given in [link] to be 50 and $70\ \mu\text{Ci}$, respectively. Find and compare the masses of ^{131}I and ^{123}I in such scans, given their respective half-lives are 8.04 d and 13.2 h. The masses are so small that the radioiodine is usually mixed with stable iodine as a carrier to ensure normal chemistry and distribution in the body.

Exercise:

Problem:

(a) Neutron activation of sodium, which is 100% ^{23}Na, produces ^{24}Na, which is used in some heart scans, as seen in [link]. The equation for the reaction is $^{23}\text{Na} + n \rightarrow {}^{24}\text{Na} + \gamma$. Find its energy output, given the mass of ^{24}Na is 23.990962 u.

(b) What mass of ^{24}Na produces the needed 5.0-mCi activity, given its half-life is 15.0 h?

Solution:

(a) 6.958 MeV

(b) 5.7×10^{-10} g

Glossary

Anger camera
: a common medical imaging device that uses a scintillator connected to a series of photomultipliers

gamma camera
: another name for an Anger camera

positron emission tomography (PET)
: tomography technique that uses β^{+} emitters and detects the two annihilation γ rays, aiding in source localization

radiopharmaceutical
: compound used for medical imaging

single-photon-emission computed tomography (SPECT)
: tomography performed with γ-emitting radiopharmaceuticals

tagged
: process of attaching a radioactive substance to a chemical compound

Biological Effects of Ionizing Radiation

- Define various units of radiation.
- Describe RBE.

We hear many seemingly contradictory things about the biological effects of ionizing radiation. It can cause cancer, burns, and hair loss, yet it is used to treat and even cure cancer. How do we understand these effects? Once again, there is an underlying simplicity in nature, even in complicated biological organisms. All the effects of ionizing radiation on biological tissue can be understood by knowing that **ionizing radiation affects molecules within cells, particularly DNA molecules.**

Let us take a brief look at molecules within cells and how cells operate. Cells have long, double-helical DNA molecules containing chemical codes called genetic codes that govern the function and processes undertaken by the cell. It is for unraveling the double-helical structure of DNA that James Watson, Francis Crick, and Maurice Wilkins received the Nobel Prize. Damage to DNA consists of breaks in chemical bonds or other changes in the structural features of the DNA chain, leading to changes in the genetic code. In human cells, we can have as many as a million individual instances of damage to DNA per cell per day. It is remarkable that DNA contains codes that check whether the DNA is damaged or can repair itself. It is like an auto check and repair mechanism. This repair ability of DNA is vital for maintaining the integrity of the genetic code and for the normal functioning of the entire organism. It should be constantly active and needs to respond rapidly. The rate of DNA repair depends on various factors such as the cell type and age of the cell. A cell with a damaged ability to repair DNA, which could have been induced by ionizing radiation, can do one of the following:

- The cell can go into an irreversible state of dormancy, known as senescence.
- The cell can commit suicide, known as programmed cell death.
- The cell can go into unregulated cell division leading to tumors and cancers.

Since ionizing radiation damages the DNA, which is critical in cell reproduction, it has its greatest effect on cells that rapidly reproduce, including most types of cancer. Thus, cancer cells are more sensitive to radiation than normal cells and can be killed by it easily. Cancer is characterized by a malfunction of cell reproduction, and can also be caused by ionizing radiation. Without contradiction, ionizing radiation can be both a cure and a cause.

To discuss quantitatively the biological effects of ionizing radiation, we need a radiation dose unit that is directly related to those effects. All effects of radiation are assumed to be directly proportional to the amount of ionization produced in the biological organism. The amount of ionization is in turn proportional to the amount of deposited energy. Therefore, we define a **radiation dose unit** called the **rad**, as $1/100$ of a joule of ionizing energy deposited per kilogram of tissue, which is

Equation:

$$1 \text{ rad} = 0.01 \text{ J/kg}.$$

For example, if a 50.0-kg person is exposed to ionizing radiation over her entire body and she absorbs 1.00 J, then her whole-body radiation dose is

Equation:

$$(1.00 \text{ J})/(50.0 \text{ kg}) = 0.0200 \text{ J/kg} = 2.00 \text{ rad}.$$

If the same 1.00 J of ionizing energy were absorbed in her 2.00-kg forearm alone, then the dose to the forearm would be

Equation:

$$(1.00 \text{ J})/(2.00 \text{ kg}) = 0.500 \text{ J/kg} = 50.0 \text{ rad},$$

and the unaffected tissue would have a zero rad dose. While calculating radiation doses, you divide the energy absorbed by the mass of affected tissue. You must specify the affected region, such as the whole body or forearm in addition to giving the numerical dose in rads. The SI unit for radiation dose is the **gray (Gy)**, which is defined to be

Equation:

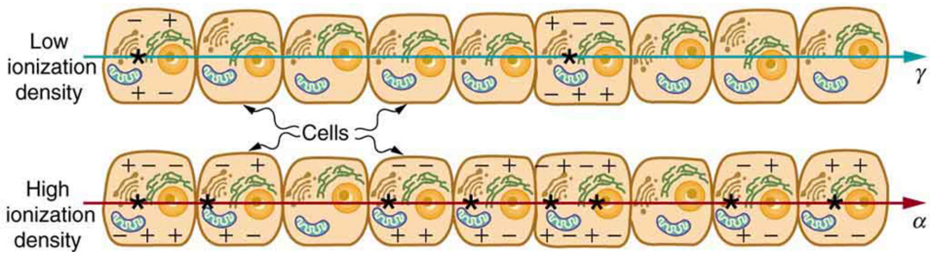
$$1 \text{ Gy} = 1 \text{ J/kg} = 100 \text{ rad}.$$

However, the rad is still commonly used. Although the energy per kilogram in 1 rad is small, it has significant effects since the energy causes ionization. The energy needed for a single ionization is a few eV, or less than 10^{-18} J. Thus, 0.01 J of ionizing energy can create a huge number of ion pairs and have an effect at the cellular level.

The effects of ionizing radiation may be directly proportional to the dose in rads, but they also depend on the type of radiation and the type of tissue. That is, for a given dose in rads, the effects depend on whether the radiation is α, β, γ, x-ray, or some other type of ionizing radiation. In the earlier discussion of the range of ionizing radiation, it was noted that energy is deposited in a series of ionizations and not in a single interaction. Each ion pair or ionization requires a certain amount of energy, so that the number of ion pairs is directly proportional to the amount of the deposited ionizing energy. But, if the range of the radiation is small, as it is for α s, then the ionization and the damage created is more concentrated and harder for the organism to repair, as seen in [link]. Concentrated damage is more difficult for biological organisms to repair than damage that is spread out, so short-range particles have greater biological effects. The **relative biological effectiveness** (RBE) or **quality factor** (QF) is given in [link] for several types of ionizing radiation—the effect of the radiation is directly proportional to the RBE. A dose unit more closely related to effects in biological tissue is called the **roentgen equivalent man** or rem and is defined to be the dose in rads multiplied by the relative biological effectiveness.

Equation:

$$\text{rem} = \text{rad} \times \text{RBE}$$



The image shows ionization created in cells by α and γ radiation. Because of its shorter range, the ionization and damage created by α is more concentrated and harder for the organism to repair. Thus, the RBE for α s is greater than the RBE for γ s, even though they create the same amount of ionization at the same energy.

So, if a person had a whole-body dose of 2.00 rad of γ radiation, the dose in rem would be $(2.00 \text{ rad})(1) = 2.00 \text{ rem whole body}$. If the person had a whole-body dose of 2.00 rad of α radiation, then the dose in rem would be $(2.00 \text{ rad})(20) = 40.0 \text{ rem whole body}$. The α s would have 20 times the effect on the person than the γ s for the same deposited energy. The SI equivalent of the rem is the **sievert** (Sv), defined to be $\text{Sv} = \text{Gy} \times \text{RBE}$, so that

Equation:

$$1 \text{ Sv} = 1 \text{ Gy} \times \text{RBE} = 100 \text{ rem}.$$

The RBEs given in [link] are approximate, but they yield certain insights. For example, the eyes are more sensitive to radiation, because the cells of the lens do not repair themselves. Neutrons cause more damage than γ rays, although both are neutral and have large ranges, because neutrons often cause secondary radiation when they are captured. Note that the RBEs are 1 for higher-energy β s, γ s, and x-rays, three of the most common types of radiation. For those types of radiation, the numerical values of the dose in rem and rad are identical. For example, 1 rad of γ radiation is also 1 rem. For that reason, rads are still widely quoted rather than rem. [link] summarizes the units that are used for radiation.

Note:

Misconception Alert: Activity vs. Dose
“Activity” refers to the radioactive source while “dose” refers to the amount of energy from the radiation that is deposited in a person or object.

A high level of activity doesn’t mean much if a person is far away from the source. The activity R of a source depends upon the quantity of material (kg) as well as the half-life. A short half-life will produce many more disintegrations per second. Recall that $R = \frac{0.693N}{t_{1/2}}$. Also, the activity decreases exponentially, which is seen in the equation $R = R_0 e^{-\lambda t}$.

Type and energy of radiation	**RBE**[footnote] **Values approximate, difficult to determine.**
X-rays	1
γ rays	1
β rays greater than 32 keV	1
βrays less than 32 keV	1.7
Neutrons, thermal to slow (<20 keV)	2–5
Neutrons, fast (1–10 MeV)	10 (body), 32 (eyes)
Protons (1–10 MeV)	10 (body), 32 (eyes)
α rays from radioactive decay	10–20
Heavy ions from accelerators	10–20

Relative Biological Effectiveness

Quantity	SI unit name	Definition	Former unit	Conversion
Activity	Becquerel (bq)	decay/sec	Curie (Ci)	$1 \text{ Bq} = 2.7 \times 10^{-11} \text{ Ci}$
Absorbed dose	Gray (Gy)	1 J/kg	rad	$\text{Gy} = 100 \text{ rad}$
Dose Equivalent	Sievert (Sv)	1 J/kg × RBE	rem	$\text{Sv} = 100 \text{ rem}$

Units for Radiation

The large-scale effects of radiation on humans can be divided into two categories: immediate effects and long-term effects. [link] gives the immediate effects of whole-body exposures received in less than one day. If the radiation exposure is spread out over more time, greater doses are needed to cause the effects listed. This is due to the body's ability to partially repair the damage. Any dose less than 100 mSv (10 rem) is called a **low dose**, 0.1 Sv to 1 Sv (10 to 100 rem) is called a **moderate dose**, and anything greater than 1 Sv (100 rem) is called a **high dose**. There is no known way to determine after the fact if a person has been exposed to less than 10 mSv.

Dose in Sv [footnote] Multiply by 100 to obtain dose in rem.	Effect
0–0.10	No observable effect.
0.1 – 1	Slight to moderate decrease in white blood cell counts.
0.5	Temporary sterility; 0.35 for women, 0.50 for men.
1 – 2	Significant reduction in blood cell counts, brief nausea and vomiting. Rarely fatal.
2 – 5	Nausea, vomiting, hair loss, severe blood damage, hemorrhage, fatalities.

Dose in Sv [footnote] Multiply by 100 to obtain dose in rem.	**Effect**
4.5	LD50/32. Lethal to 50% of the population within 32 days after exposure if not treated.
5 – 20	Worst effects due to malfunction of small intestine and blood systems. Limited survival.
>20	Fatal within hours due to collapse of central nervous system.

Immediate Effects of Radiation (Adults, Whole Body, Single Exposure)

Immediate effects are explained by the effects of radiation on cells and the sensitivity of rapidly reproducing cells to radiation. The first clue that a person has been exposed to radiation is a change in blood count, which is not surprising since blood cells are the most rapidly reproducing cells in the body. At higher doses, nausea and hair loss are observed, which may be due to interference with cell reproduction. Cells in the lining of the digestive system also rapidly reproduce, and their destruction causes nausea. When the growth of hair cells slows, the hair follicles become thin and break off. High doses cause significant cell death in all systems, but the lowest doses that cause fatalities do so by weakening the immune system through the loss of white blood cells.

The two known long-term effects of radiation are cancer and genetic defects. Both are directly attributable to the interference of radiation with cell reproduction. For high doses of radiation, the risk of cancer is reasonably well known from studies of exposed groups. Hiroshima and Nagasaki survivors and a smaller number of people exposed by their occupation, such as radium dial painters, have been fully documented. Chernobyl victims will be studied for many decades, with some data already available. For example, a significant increase in childhood thyroid cancer has been observed. The risk of a radiation-induced cancer for low and moderate doses is generally *assumed* to be proportional to the risk known for high doses. Under this assumption, any dose of radiation, no matter how small, involves a risk to human health. This is called the **linear hypothesis** and it may be prudent, but it *is* controversial. There is some evidence that, unlike the immediate effects of radiation, the long-term effects are cumulative and there is little self-repair. This is analogous to the risk of skin cancer from UV exposure, which is known to be cumulative.

There is a latency period for the onset of radiation-induced cancer of about 2 years for leukemia and 15 years for most other forms. The person is at risk for at least 30 years after the latency period. Omitting many details, the overall risk of a radiation-induced cancer

death per year per rem of exposure is about 10 in a million, which can be written as $10/10^6\ \text{rem}\cdot\text{y}$.

If a person receives a dose of 1 rem, his risk each year of dying from radiation-induced cancer is 10 in a million and that risk continues for about 30 years. The lifetime risk is thus 300 in a million, or 0.03 percent. Since about 20 percent of all worldwide deaths are from cancer, the increase due to a 1 rem exposure is impossible to detect demographically. But 100 rem (1 Sv), which was the dose received by the average Hiroshima and Nagasaki survivor, causes a 3 percent risk, which can be observed in the presence of a 20 percent normal or natural incidence rate.

The incidence of genetic defects induced by radiation is about one-third that of cancer deaths, but is much more poorly known. The lifetime risk of a genetic defect due to a 1 rem exposure is about 100 in a million or $3.3/10^6\ \text{rem}\cdot\text{y}$, but the normal incidence is 60,000 in a million. Evidence of such a small increase, tragic as it is, is nearly impossible to obtain. For example, there is no evidence of increased genetic defects among the offspring of Hiroshima and Nagasaki survivors. Animal studies do not seem to correlate well with effects on humans and are not very helpful. For both cancer and genetic defects, the approach to safety has been to use the linear hypothesis, which is likely to be an overestimate of the risks of low doses. Certain researchers even claim that low doses are *beneficial*. **Hormesis** is a term used to describe generally favorable biological responses to low exposures of toxins or radiation. Such low levels may help certain repair mechanisms to develop or enable cells to adapt to the effects of the low exposures. Positive effects may occur at low doses that could be a problem at high doses.

Even the linear hypothesis estimates of the risks are relatively small, and the average person is not exposed to large amounts of radiation. [link] lists average annual background radiation doses from natural and artificial sources for Australia, the United States, Germany, and world-wide averages. Cosmic rays are partially shielded by the atmosphere, and the dose depends upon altitude and latitude, but the average is about 0.40 mSv/y. A good example of the variation of cosmic radiation dose with altitude comes from the airline industry. Monitored personnel show an average of 2 mSv/y. A 12-hour flight might give you an exposure of 0.02 to 0.03 mSv.

Doses from the Earth itself are mainly due to the isotopes of uranium, thorium, and potassium, and vary greatly by location. Some places have great natural concentrations of uranium and thorium, yielding doses ten times as high as the average value. Internal doses come from foods and liquids that we ingest. Fertilizers containing phosphates have potassium and uranium. So we are all a little radioactive. Carbon-14 has about 66 Bq/kg radioactivity whereas fertilizers may have more than 3000 Bq/kg radioactivity. Medical and dental diagnostic exposures are mostly from x-rays. It should be noted that x-ray doses tend to be localized and are becoming much smaller with improved techniques. [link] shows typical doses received during various diagnostic x-ray examinations. Note the large dose from a CT scan. While CT scans only account for less than 20 percent of

the x-ray procedures done today, they account for about 50 percent of the annual dose received.

Radon is usually more pronounced underground and in buildings with low air exchange with the outside world. Almost all soil contains some ^{226}Ra and ^{222}Rn, but radon is lower in mainly sedimentary soils and higher in granite soils. Thus, the exposure to the public can vary greatly, even within short distances. Radon can diffuse from the soil into homes, especially basements. The estimated exposure for ^{222}Rn is controversial. Recent studies indicate there is more radon in homes than had been realized, and it is speculated that radon may be responsible for 20 percent of lung cancers, being particularly hazardous to those who also smoke. Many countries have introduced limits on allowable radon concentrations in indoor air, often requiring the measurement of radon concentrations in a house prior to its sale. Ironically, it could be argued that the higher levels of radon exposure and their geographic variability, taken with the lack of demographic evidence of any effects, means that low-level radiation is *less* dangerous than previously thought.

Radiation Protection

Laws regulate radiation doses to which people can be exposed. The greatest occupational whole-body dose that is allowed depends upon the country and is about 20 to 50 mSv/y and is rarely reached by medical and nuclear power workers. Higher doses are allowed for the hands. Much lower doses are permitted for the reproductive organs and the fetuses of pregnant women. Inadvertent doses to the public are limited to $1/10$ of occupational doses, except for those caused by nuclear power, which cannot legally expose the public to more than $1/1000$ of the occupational limit or 0.05 mSv/y (5 mrem/y). This has been exceeded in the United States only at the time of the Three Mile Island (TMI) accident in 1979. Chernobyl is another story. Extensive monitoring with a variety of radiation detectors is performed to assure radiation safety. Increased ventilation in uranium mines has lowered the dose there to about 1 mSv/y.

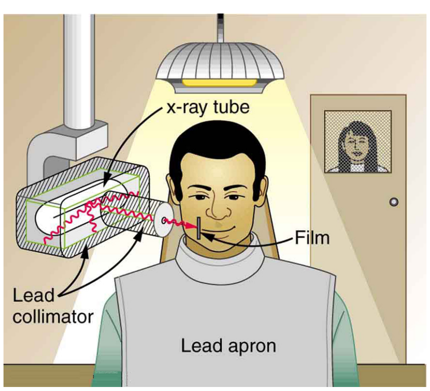
Source	**Dose (mSv/y)[footnote]** **Multiply by 100 to obtain dose in mrem/y.**			
Source	Australia	Germany	United States	World
Natural Radiation - external				

Source	Dose (mSv/y)[footnote] Multiply by 100 to obtain dose in mrem/y.			
Cosmic Rays	0.30	0.28	0.30	0.39
Soil, building materials	0.40	0.40	0.30	0.48
Radon gas	0.90	1.1	2.0	1.2
Natural Radiation - internal				
^{40}K, ^{14}C, ^{226}Ra	0.24	0.28	0.40	0.29
Medical & Dental	0.80	0.90	0.53	0.40
TOTAL	2.6	3.0	3.5	2.8

Background Radiation Sources and Average Doses

To physically limit radiation doses, we use **shielding**, increase the **distance** from a source, and limit the **time of exposure**.

[link] illustrates how these are used to protect both the patient and the dental technician when an x-ray is taken. Shielding absorbs radiation and can be provided by any material, including sufficient air. The greater the distance from the source, the more the radiation spreads out. The less time a person is exposed to a given source, the smaller is the dose received by the person. Doses from most medical diagnostics have decreased in recent years due to faster films that require less exposure time.



A lead apron is placed over the dental patient and shielding surrounds the x-ray tube to limit exposure to tissue other than the tissue that is being imaged. Fast films limit the time needed to obtain images, reducing exposure to the imaged tissue. The technician stands a few meters away behind a lead-lined door with a lead glass window, reducing her occupational exposure.

Procedure	**Effective dose (mSv)**
Chest	0.02
Dental	0.01
Skull	0.07
Leg	0.02
Mammogram	0.40
Barium enema	7.0
Upper GI	3.0
CT head	2.0
CT abdomen	10.0

Typical Doses Received During Diagnostic X-ray Exams

Problem-Solving Strategy

You need to follow certain steps for dose calculations, which are

Step 1. *Examine the situation to determine that a person is exposed to ionizing radiation.*

Step 2. *Identify exactly what needs to be determined in the problem (identify the unknowns).* The most straightforward problems ask for a dose calculation.

Step 3. *Make a list of what is given or can be inferred from the problem as stated (identify the knowns).* Look for information on the type of radiation, the energy per event, the activity, and the mass of tissue affected.

Step 4. *For dose calculations, you need to determine the energy deposited.* This may take one or more steps, depending on the given information.

Step 5. *Divide the deposited energy by the mass of the affected tissue.* Use units of joules for energy and kilograms for mass. If a dose in Sv is involved, use the definition that $1\ \text{Sv} = 1\ \text{J/kg}$.

Step 6. *If a dose in mSv is involved, determine the RBE (QF) of the radiation.* Recall that $1\ \text{mSv} = 1\ \text{mGy} \times \text{RBE}\ (\text{or}\ 1\ \text{rem} = 1\ \text{rad} \times \text{RBE})$.

Step 7. *Check the answer to see if it is reasonable: Does it make sense?* The dose should be consistent with the numbers given in the text for diagnostic, occupational, and therapeutic exposures.

Example:
Dose from Inhaled Plutonium
Calculate the dose in rem/y for the lungs of a weapons plant employee who inhales and retains an activity of 1.00 μCi of ^{239}Pu in an accident. The mass of affected lung tissue is 2.00 kg, the plutonium decays by emission of a 5.23-MeV α particle, and you may assume the higher value of the RBE for α s from [link].
Strategy
Dose in rem is defined by $1\ \text{rad} = 0.01\ \text{J/kg}$ and $\text{rem} = \text{rad} \times \text{RBE}$. The energy deposited is divided by the mass of tissue affected and then multiplied by the RBE. The latter two quantities are given, and so the main task in this example will be to find the energy deposited in one year. Since the activity of the source is given, we can calculate the number of decays, multiply by the energy per decay, and convert MeV to joules to get the total energy.
Solution

The activity $R = 1.00\ \mu\text{Ci} = 3.70 \times 10^4\ \text{Bq} = 3.70 \times 10^4$ decays/s. So, the number of decays per year is obtained by multiplying by the number of seconds in a year:

Equation:

$$\left(3.70 \times 10^4\ \text{decays/s}\right)\left(3.16 \times 10^7\ \text{s}\right) = 1.17 \times 10^{12}\ \text{decays.}$$

Thus, the ionizing energy deposited per year is

Equation:

$$E = \left(1.17 \times 10^{12}\ \text{decays}\right)(5.23\ \text{MeV/decay}) \times \left(\frac{1.60 \times 10^{-13}\ \text{J}}{\text{MeV}}\right) = 0.978\ \text{J.}$$

Dividing by the mass of the affected tissue gives

Equation:

$$\frac{E}{\text{mass}} = \frac{0.978\ \text{J}}{2.00\ \text{kg}} = 0.489\ \text{J/kg.}$$

One Gray is 1.00 J/kg, and so the dose in Gy is

Equation:

$$\text{dose in Gy} = \frac{0.489\ \text{J/kg}}{1.00\ (\text{J/kg})/\text{Gy}} = 0.489\ \text{Gy.}$$

Now, the dose in Sv is

Equation:

$$\text{dose in Sv} = \text{Gy} \times \text{RBE}$$

Equation:

$$= (0.489\ \text{Gy})(20) = 9.8\ \text{Sv.}$$

Discussion

First note that the dose is given to two digits, because the RBE is (at best) known only to two digits. By any standard, this yearly radiation dose is high and will have a devastating effect on the health of the worker. Worse yet, plutonium has a long radioactive half-life and is not readily eliminated by the body, and so it will remain in the lungs. Being an α emitter makes the effects 10 to 20 times worse than the same ionization produced by β s, γ rays, or x-rays. An activity of 1.00 μCi is created by only 16 μg of ^{239}Pu (left as an end-of-chapter problem to verify), partly justifying claims that plutonium is the most toxic substance known. Its actual hazard depends on how likely it is to be spread out among a large population and then ingested. The Chernobyl disaster's deadly legacy, for example, has nothing to do with the plutonium it put into the environment.

Risk versus Benefit

Medical doses of radiation are also limited. Diagnostic doses are generally low and have further lowered with improved techniques and faster films. With the possible exception of routine dental x-rays, radiation is used diagnostically only when needed so that the low risk is justified by the benefit of the diagnosis. Chest x-rays give the lowest doses—about 0.1 mSv to the tissue affected, with less than 5 percent scattering into tissues that are not directly imaged. Other x-ray procedures range upward to about 10 mSv in a CT scan, and about 5 mSv (0.5 rem) per dental x-ray, again both only affecting the tissue imaged. Medical images with radiopharmaceuticals give doses ranging from 1 to 5 mSv, usually localized. One exception is the thyroid scan using ^{131}I. Because of its relatively long half-life, it exposes the thyroid to about 0.75 Sv. The isotope ^{123}I is more difficult to produce, but its short half-life limits thyroid exposure to about 15 mSv.

Note:
PhET Explorations: Alpha Decay
Watch alpha particles escape from a polonium nucleus, causing radioactive alpha decay. See how random decay times relate to the half life.

[Alpha Decay]

Section Summary

- The biological effects of ionizing radiation are due to two effects it has on cells: interference with cell reproduction, and destruction of cell function.
- A radiation dose unit called the rad is defined in terms of the ionizing energy deposited per kilogram of tissue:
  Equation:
  $$1\ \text{rad} = 0.01\ \text{J/kg}.$$
- The SI unit for radiation dose is the gray (Gy), which is defined to be $1\ \text{Gy} = 1\ \text{J/kg} = 100\ \text{rad}.$
- To account for the effect of the type of particle creating the ionization, we use the relative biological effectiveness (RBE) or quality factor (QF) given in [link] and define a unit called the roentgen equivalent man (rem) as

Equation:

$$\text{rem} = \text{rad} \times \text{RBE}.$$

- Particles that have short ranges or create large ionization densities have RBEs greater than unity. The SI equivalent of the rem is the sievert (Sv), defined to be
  Equation:

$$\text{Sv} = \text{Gy} \times \text{RBE and } 1\text{ Sv} = 100\text{ rem}.$$

- Whole-body, single-exposure doses of 0.1 Sv or less are low doses while those of 0.1 to 1 Sv are moderate, and those over 1 Sv are high doses. Some immediate radiation effects are given in [link]. Effects due to low doses are not observed, but their risk is assumed to be directly proportional to those of high doses, an assumption known as the linear hypothesis. Long-term effects are cancer deaths at the rate of $10/10^6\text{ rem}\cdot\text{y}$and genetic defects at roughly one-third this rate. Background radiation doses and sources are given in [link]. World-wide average radiation exposure from natural sources, including radon, is about 3 mSv, or 300 mrem. Radiation protection utilizes shielding, distance, and time to limit exposure.

Conceptual Questions

Exercise:

Problem:

Isotopes that emit α radiation are relatively safe outside the body and exceptionally hazardous inside. Yet those that emit γ radiation are hazardous outside and inside. Explain why.

Exercise:

Problem:

Why is radon more closely associated with inducing lung cancer than other types of cancer?

Exercise:

Problem:

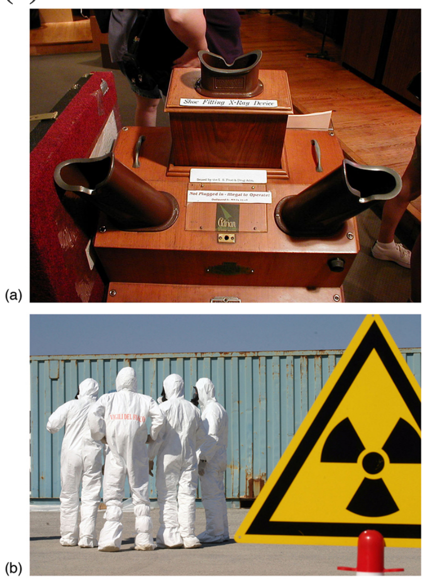
The RBE for low-energy βs is 1.7, whereas that for higher-energy βs is only 1. Explain why, considering how the range of radiation depends on its energy.

Exercise:

Problem:

Which methods of radiation protection were used in the device shown in the first photo in [link]? Which were used in the situation shown in the second photo?

(a)



(a) This x-ray fluorescence machine is one of the thousands used in shoe stores to produce images of feet as a check on the fit of shoes. They are unshielded and remain on as long as the feet are in them, producing doses much greater than medical images. Children were fascinated with them. These machines were used in shoe stores until laws preventing such unwarranted radiation exposure were enacted in the 1950s. (credit: Andrew Kuchling) (b) Now that we know the effects of exposure to

radioactive material, safety is a priority. (credit: U.S. Navy)

Exercise:

Problem:

What radioisotope could be a problem in homes built of cinder blocks made from uranium mine tailings? (This is true of homes and schools in certain regions near uranium mines.)

Exercise:

Problem:

Are some types of cancer more sensitive to radiation than others? If so, what makes them more sensitive?

Exercise:

Problem:

Suppose a person swallows some radioactive material by accident. What information is needed to be able to assess possible damage?

Problems & Exercises

Exercise:

Problem:

What is the dose in mSv for: (a) a 0.1 Gy x-ray? (b) 2.5 mGy of neutron exposure to the eye? (c) 1.5 mGy of α exposure?

Solution:

(a) 100 mSv

(b) 80 mSv

(c) ~30 mSv

Exercise:

Problem:

Find the radiation dose in Gy for: (a) A 10-mSv fluoroscopic x-ray series. (b) 50 mSv of skin exposure by an α emitter. (c) 160 mSv of β^- and γ rays from the ^{40}K in your body.

Exercise:

Problem:

How many Gy of exposure is needed to give a cancerous tumor a dose of 40 Sv if it is exposed to α activity?

Solution:

~2 Gy

Exercise:

Problem:

What is the dose in Sv in a cancer treatment that exposes the patient to 200 Gy of γ rays?

Exercise:

Problem:

One half the γ rays from ^{99m}Tc are absorbed by a 0.170-mm-thick lead shielding. Half of the γ rays that pass through the first layer of lead are absorbed in a second layer of equal thickness. What thickness of lead will absorb all but one in 1000 of these γ rays?

Solution:

1.69 mm

Exercise:

Problem:

A plumber at a nuclear power plant receives a whole-body dose of 30 mSv in 15 minutes while repairing a crucial valve. Find the radiation-induced yearly risk of death from cancer and the chance of genetic defect from this maximum allowable exposure.

Exercise:

Problem:

In the 1980s, the term picowave was used to describe food irradiation in order to overcome public resistance by playing on the well-known safety of microwave radiation. Find the energy in MeV of a photon having a wavelength of a picometer.

Solution:

1.24 MeV

Exercise:

Problem: Find the mass of ^{239}Pu that has an activity of 1.00 μCi.

Glossary

gray (Gy)
: the SI unit for radiation dose which is defined to be $1 \text{ Gy} = 1 \text{ J/kg} = 100 \text{ rad}$

linear hypothesis
: assumption that risk is directly proportional to risk from high doses

rad
: the ionizing energy deposited per kilogram of tissue

sievert
: the SI equivalent of the rem

relative biological effectiveness (RBE)
: a number that expresses the relative amount of damage that a fixed amount of ionizing radiation of a given type can inflict on biological tissues

quality factor
: same as relative biological effectiveness

roentgen equivalent man (rem)
: a dose unit more closely related to effects in biological tissue

low dose
: a dose less than 100 mSv (10 rem)

moderate dose
: a dose from 0.1 Sv to 1 Sv (10 to 100 rem)

high dose

a dose greater than 1 Sv (100 rem)

hormesis

a term used to describe generally favorable biological responses to low exposures of toxins or radiation

shielding

a technique to limit radiation exposure

Therapeutic Uses of Ionizing Radiation

- Explain the concept of radiotherapy and list typical doses for cancer therapy.

Therapeutic applications of ionizing radiation, called radiation therapy or **radiotherapy**, have existed since the discovery of x-rays and nuclear radioactivity. Today, radiotherapy is used almost exclusively for cancer therapy, where it saves thousands of lives and improves the quality of life and longevity of many it cannot save. Radiotherapy may be used alone or in combination with surgery and chemotherapy (drug treatment) depending on the type of cancer and the response of the patient. A careful examination of all available data has established that radiotherapy's beneficial effects far outweigh its long-term risks.

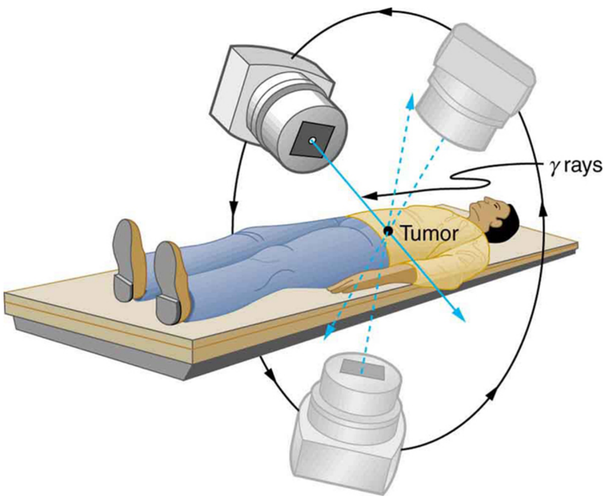
Medical Application

The earliest uses of ionizing radiation on humans were mostly harmful, with many at the level of snake oil as seen in [link]. Radium-doped cosmetics that glowed in the dark were used around the time of World War I. As recently as the 1950s, radon mine tours were promoted as healthful and rejuvenating—those who toured were exposed but gained no benefits. Radium salts were sold as health elixirs for many years. The gruesome death of a wealthy industrialist, who became psychologically addicted to the brew, alerted the unsuspecting to the dangers of radium salt elixirs. Most abuses finally ended after the legislation in the 1950s.

The properties of radiation were once touted for far more than its modern use in cancer therapy. Until 1932, radium was advertised for a variety of uses, often with tragic results. (credit: Struthious Bandersnatch.)

Radiotherapy is effective against cancer because cancer cells reproduce rapidly and, consequently, are more sensitive to radiation. The central problem in radiotherapy is to make the dose for cancer cells as high as possible while limiting the dose for normal cells. The ratio of abnormal cells killed to normal cells killed is called the **therapeutic ratio**, and all radiotherapy techniques are designed to enhance this ratio. Radiation can be concentrated in cancerous tissue by a number of techniques. One of the most prevalent techniques for well-defined tumors is a geometric technique

shown in [link]. A narrow beam of radiation is passed through the patient from a variety of directions with a common crossing point in the tumor. This concentrates the dose in the tumor while spreading it out over a large volume of normal tissue. The external radiation can be x-rays, ^{60}Co γ rays, or ionizing-particle beams produced by accelerators. Accelerator-produced beams of neutrons, π-mesons, and heavy ions such as nitrogen nuclei have been employed, and these can be quite effective. These particles have larger QFs or RBEs and sometimes can be better localized, producing a greater therapeutic ratio. But accelerator radiotherapy is much more expensive and less frequently employed than other forms.



The ^{60}Co source of γ-radiation is rotated around the patient so that the common crossing point is in the tumor, concentrating the dose there. This geometric technique works for well-defined tumors.

Another form of radiotherapy uses chemically inert radioactive implants. One use is for prostate cancer. Radioactive seeds (about 40 to 100 and the size of a grain of rice) are placed in the prostate region. The isotopes used

are usually ^{135}I (6-month half life) or ^{103}Pd (3-month half life). Alpha emitters have the dual advantages of a large QF and a small range for better localization.

Radiopharmaceuticals are used for cancer therapy when they can be localized well enough to produce a favorable therapeutic ratio. Thyroid cancer is commonly treated utilizing radioactive iodine. Thyroid cells concentrate iodine, and cancerous thyroid cells are more aggressive in doing this. An ingenious use of radiopharmaceuticals in cancer therapy tags antibodies with radioisotopes. Antibodies produced by a patient to combat his cancer are extracted, cultured, loaded with a radioisotope, and then returned to the patient. The antibodies are concentrated almost entirely in the tissue they developed to fight, thus localizing the radiation in abnormal tissue. The therapeutic ratio can be quite high for short-range radiation. There is, however, a significant dose for organs that eliminate radiopharmaceuticals from the body, such as the liver, kidneys, and bladder. As with most radiotherapy, the technique is limited by the tolerable amount of damage to the normal tissue.

[link] lists typical therapeutic doses of radiation used against certain cancers. The doses are large, but not fatal because they are localized and spread out in time. Protocols for treatment vary with the type of cancer and the condition and response of the patient. Three to five 200-rem treatments per week for a period of several weeks is typical. Time between treatments allows the body to repair normal tissue. This effect occurs because damage is concentrated in the abnormal tissue, and the abnormal tissue is more sensitive to radiation. Damage to normal tissue limits the doses. You will note that the greatest doses are given to any tissue that is not rapidly reproducing, such as in the adult brain. Lung cancer, on the other end of the scale, cannot ordinarily be cured with radiation because of the sensitivity of lung tissue and blood to radiation. But radiotherapy for lung cancer does alleviate symptoms and prolong life and is therefore justified in some cases.

Type of Cancer	Typical dose (Sv)
Lung	10–20
Hodgkin's disease	40–45
Skin	40–50
Ovarian	50–75
Breast	50–80+
Brain	80+
Neck	80+
Bone	80+
Soft tissue	80+
Thyroid	80+

Cancer Radiotherapy

Finally, it is interesting to note that chemotherapy employs drugs that interfere with cell division and is, thus, also effective against cancer. It also has almost the same side effects, such as nausea and hair loss, and risks, such as the inducement of another cancer.

Section Summary

- Radiotherapy is the use of ionizing radiation to treat ailments, now limited to cancer therapy.
- The sensitivity of cancer cells to radiation enhances the ratio of cancer cells killed to normal cells killed, which is called the therapeutic ratio.

- Doses for various organs are limited by the tolerance of normal tissue for radiation. Treatment is localized in one region of the body and spread out in time.

Conceptual Questions

Exercise:

Problem:

Radiotherapy is more likely to be used to treat cancer in elderly patients than in young ones. Explain why. Why is radiotherapy used to treat young people at all?

Problems & Exercises

Exercise:

Problem:

A beam of 168-MeV nitrogen nuclei is used for cancer therapy. If this beam is directed onto a 0.200-kg tumor and gives it a 2.00-Sv dose, how many nitrogen nuclei were stopped? (Use an RBE of 20 for heavy ions.)

Solution:

7.44×10^8

Exercise:

Problem:

(a) If the average molecular mass of compounds in food is 50.0 g, how many molecules are there in 1.00 kg of food? (b) How many ion pairs are created in 1.00 kg of food, if it is exposed to 1000 Sv and it takes 32.0 eV to create an ion pair? (c) Find the ratio of ion pairs to molecules. (d) If these ion pairs recombine into a distribution of 2000 new compounds, how many parts per billion is each?

Exercise:

Problem:

Calculate the dose in Sv to the chest of a patient given an x-ray under the following conditions. The x-ray beam intensity is $1.50\ \mathrm{W/m^2}$, the area of the chest exposed is $0.0750\ \mathrm{m^2}$, 35.0% of the x-rays are absorbed in 20.0 kg of tissue, and the exposure time is 0.250 s.

Solution:

4.92×10^{-4} Sv

Exercise:

Problem:

(a) A cancer patient is exposed to γ rays from a 5000-Ci $^{60}\mathrm{Co}$ transillumination unit for 32.0 s. The γ rays are collimated in such a manner that only 1.00% of them strike the patient. Of those, 20.0% are absorbed in a tumor having a mass of 1.50 kg. What is the dose in rem to the tumor, if the average γ energy per decay is 1.25 MeV? None of the β s from the decay reach the patient. (b) Is the dose consistent with stated therapeutic doses?

Exercise:

Problem:

What is the mass of $^{60}\mathrm{Co}$ in a cancer therapy transillumination unit containing 5.00 kCi of $^{60}\mathrm{Co}$?

Solution:

4.43 g

Exercise:

Problem:

Large amounts of $^{65}\mathrm{Zn}$ are produced in copper exposed to accelerator beams. While machining contaminated copper, a physicist ingests $50.0\ \mu\mathrm{Ci}$ of $^{65}\mathrm{Zn}$. Each $^{65}\mathrm{Zn}$ decay emits an average γ-ray energy of 0.550 MeV, 40.0% of which is absorbed in the scientist's 75.0-kg body. What dose in mSv is caused by this in one day?

Exercise:

Problem:

Naturally occurring $^{40}\mathrm{K}$ is listed as responsible for 16 mrem/y of background radiation. Calculate the mass of $^{40}\mathrm{K}$ that must be inside the 55-kg body of a woman to produce this dose. Each $^{40}\mathrm{K}$ decay emits a 1.32-MeV β, and 50% of the energy is absorbed inside the body.

Solution:

0.010 g

Exercise:

Problem:

(a) Background radiation due to $^{226}\mathrm{Ra}$ averages only 0.01 mSv/y, but it can range upward depending on where a person lives. Find the mass of $^{226}\mathrm{Ra}$ in the 80.0-kg body of a man who receives a dose of 2.50-mSv/y from it, noting that each $^{226}\mathrm{Ra}$ decay emits a 4.80-MeV α particle. You may neglect dose due to daughters and assume a constant amount, evenly distributed due to balanced ingestion and bodily elimination. (b) Is it surprising that such a small mass could cause a measurable radiation dose? Explain.

Exercise:

Problem:

The annual radiation dose from ^{14}C in our bodies is 0.01 mSv/y. Each ^{14}C decay emits a β^- averaging 0.0750 MeV. Taking the fraction of ^{14}C to be 1.3×10^{-12} N of normal ^{12}C, and assuming the body is 13% carbon, estimate the fraction of the decay energy absorbed. (The rest escapes, exposing those close to you.)

Solution:

95%

Exercise:

Problem:

If everyone in Australia received an extra 0.05 mSv per year of radiation, what would be the increase in the number of cancer deaths per year? (Assume that time had elapsed for the effects to become apparent.) Assume that there are 200×10^{-4} deaths per Sv of radiation per year. What percent of the actual number of cancer deaths recorded is this?

Glossary

radiotherapy
: the use of ionizing radiation to treat ailments

therapeutic ratio
: the ratio of abnormal cells killed to normal cells killed

Selected Radioactive Isotopes

Decay modes are α, β^-, β^+, electron capture (EC) and isomeric transition (IT). EC results in the same daughter nucleus as would β^+ decay. IT is a transition from a metastable excited state. Energies for $\beta^\pm$ decays are the maxima; average energies are roughly one-half the maxima.

Isotope	***t* 1/2**	**DecayMode(s)**	**Energy(MeV)**	**Percent**		**γ-Ray Energy(MeV)**
^{3}H	12.33 y	β^-	0.0186	100%		
^{14}C	5730 y	β^-	0.156	100%		
^{13}N	9.96 min	β^+	1.20	100%		
^{22}Na	2.602 y	β^+	0.55	90%	γ	1.27
^{32}P	14.28 d	β^-	1.71	100%		
^{35}S	87.4 d	β^-	0.167	100%		
^{36}Cl	3.00×10^5y	β^-	0.710	100%		
^{40}K	1.28×10^9y	β^-	1.31	89%		
^{43}K	22.3 h	β^-	0.827	87%	γs	0.373
						0.618
^{45}Ca	165 d	β^-	0.257	100%		
^{51}Cr	27.70 d	EC			γ	0.320
^{52}Mn	5.59d	β^+	3.69	28%	γs	1.33
						1.43
^{52}Fe	8.27 h	β^+	1.80	43%		0.169
						0.378
^{59}Fe	44.6 d	β^-s	0.273	45%	γs	1.10
			0.466	55%		1.29
^{60}Co	5.271 y	β^-	0.318	100%	γs	1.17
						1.33
^{65}Zn	244.1 d	EC			γ	1.12

Isotope	$t_{1/2}$	DecayMode(s)	Energy(MeV)	Percent		γ-Ray Energy(MeV)
^{67}Ga	78.3 h	EC			γs	0.0933
						0.185
						0.300
						others
^{75}Se	118.5 d	EC			γs	0.121
						0.136
						0.265
						0.280
						others
^{86}Rb	18.8 d	β^-s	0.69	9%	γ	1.08
			1.77	91%		
^{85}Sr	64.8 d	EC			γ	0.514
^{90}Sr	28.8 y	β^-	0.546	100%		
^{90}Y	64.1 h	β^-	2.28	100%		
^{99m}Tc	6.02 h	IT			γ	0.142
^{113m}In	99.5 min	IT			γ	0.392
^{123}I	13.0 h	EC			γ	0.159
^{131}I	8.040 d	β^-s	0.248	7%	γs	0.364
			0.607	93%		others
			others			
^{129}Cs	32.3 h	EC			γs	0.0400
						0.372
						0.411
						others
^{137}Cs	30.17 y	β^-s	0.511	95%	γ	0.662
			1.17	5%		

Isotope	$t_{1/2}$	DecayMode(s)	Energy(MeV)	Percent		γ-Ray Energy(MeV)
^{140}Ba	12.79 d	β^-	1.035	$\approx 100\%$	γs	0.030
						0.044
						0.537
						others
^{198}Au	2.696 d	β^-	1.161	$\approx 100\%$	γ	0.412
^{197}Hg	64.1 h	EC			γ	0.0733
^{210}Po	138.38 d	α	5.41	100%		
^{226}Ra	$1.60 \times 10^3 \text{y}$	αs	4.68	5%	γ	0.186
			4.87	95%		
^{235}U	$7.038 \times 10^8 \text{y}$	α	4.68	$\approx 100\%$	γs	numerous
^{238}U	$4.468 \times 10^9 \text{y}$	αs	4.22	23%	γ	0.050
			4.27	77%		
^{237}Np	$2.14 \times 10^6 \text{y}$	αs	numerous		γs	numerous
			4.96 (max.)			
^{239}Pu	$2.41 \times 10^4 \text{y}$	αs	5.19	11%	γs	7.5×10^{-5}
			5.23	15%		0.013
			5.24	73%		0.052
						others
^{243}Am	$7.37 \times 10^3 \text{y}$	αs	Max. 5.44		γs	0.075
			5.37	88%		others
			5.32	11%		
			others			

Selected Radioactive Isotopes

Glossary of Key Symbols and Notation

In this glossary, key symbols and notation are briefly defined.

Symbol	Definition
any symbol	average (indicated by a bar over a symbol—e.g., v is average velocity)
$^\circ\text{C}$	Celsius degree
$^\circ\text{F}$	Fahrenheit degree
$//$	parallel
$\perp$	perpendicular
$\propto$	proportional to
$\pm$	plus or minus

Symbol	**Definition**
0	zero as a subscript denotes an initial value
α	alpha rays
α	angular acceleration
α	temperature coefficient(s) of resistivity
β	beta rays
β	sound level
β	volume coefficient of expansion
β^-	electron emitted in nuclear beta decay
β^+	positron decay
γ	gamma rays

Symbol	Definition
γ	surface tension
$\gamma = 1/\sqrt{1 - v^2/c^2}$	a constant used in relativity
Δ	change in whatever quantity follows
δ	uncertainty in whatever quantity follows
$\varDelta E$	change in energy between the initial and final orbits of an electron in an atom
$\varDelta E$	uncertainty in energy
$\varDelta m$	difference in mass between initial and final products
$\varDelta N$	number of decays that occur
$\varDelta p$	change in momentum

Symbol	Definition
Δp	uncertainty in momentum
$\Delta \mathrm{PE}_{\mathrm{g}}$	change in gravitational potential energy
$\Delta \theta$	rotation angle
Δs	distance traveled along a circular path
Δt	uncertainty in time
Δt_0	proper time as measured by an observer at rest relative to the process
ΔV	potential difference
Δx	uncertainty in position
ε_0	permittivity of free space
η	viscosity

Symbol	Definition
θ	angle between the force vector and the displacement vector
θ	angle between two lines
θ	contact angle
θ	direction of the resultant
θ_b	Brewster's angle
θ_c	critical angle
κ	dielectric constant
λ	decay constant of a nuclide
λ	wavelength
λ_n	wavelength in a medium

Symbol	Definition
μ_0	permeability of free space
μ_{k}	coefficient of kinetic friction
μ_{s}	coefficient of static friction
v_e	electron neutrino
π^+	positive pion
π^-	negative pion
π^0	neutral pion
ρ	density
ρ_{c}	critical density, the density needed to just halt universal expansion
ρ_{fl}	fluid density

Symbol	Definition
ρ_{obj}	average density of an object
ρ/ρ_{w}	specific gravity
τ	characteristic time constant for a resistance and inductance (RL) or resistance and capacitance (RC) circuit
τ	characteristic time for a resistor and capacitor (RC) circuit
τ	torque
Υ	upsilon meson
Φ	magnetic flux
ϕ	phase angle
Ω	ohm (unit)
ω	angular velocity

Symbol	Definition
A	ampere (current unit)
A	area
A	cross-sectional area
A	total number of nucleons
a	acceleration
a_{B}	Bohr radius
a_{c}	centripetal acceleration
a_{t}	tangential acceleration
AC	alternating current
AM	amplitude modulation

Symbol	Definition
atm	atmosphere
B	baryon number
B	blue quark color
B	antiblue (yellow) antiquark color
b	quark flavor bottom or beauty
B	bulk modulus
B	magnetic field strength
$\mathrm{B}_{\mathrm{int}}$	electron’s intrinsic magnetic field
$\mathrm{B}_{\mathrm{orb}}$	orbital magnetic field
BE	binding energy of a nucleus—it is the energy required to completely disassemble it into separate protons and neutrons

Symbol	Definition
BE/A	binding energy per nucleon
Bq	becquerel—one decay per second
C	capacitance (amount of charge stored per volt)
C	coulomb (a fundamental SI unit of charge)
C_{p}	total capacitance in parallel
C_{s}	total capacitance in series
CG	center of gravity
CM	center of mass
c	quark flavor charm
c	specific heat

Symbol	Definition
c	speed of light
Cal	kilocalorie
cal	calorie
COP_{hp}	heat pump’s coefficient of performance
COP_{ref}	coefficient of performance for refrigerators and air conditioners
$\cos\theta$	cosine
$\cot\theta$	cotangent
$\csc\theta$	cosecant
D	diffusion constant
d	displacement

Symbol	Definition
d	quark flavor down
dB	decibel
d_{i}	distance of an image from the center of a lens
d_{o}	distance of an object from the center of a lens
DC	direct current
E	electric field strength
ε	emf (voltage) or Hall electromotive force
emf	electromotive force
E	energy of a single photon
E	nuclear reaction energy

Symbol	Definition
E	relativistic total energy
E	total energy
E_0	ground state energy for hydrogen
E_0	rest energy
EC	electron capture
E_{cap}	energy stored in a capacitor
Eff	efficiency—the useful work output divided by the energy input
$\mathrm{Eff}_{\mathrm{C}}$	Carnot efficiency
E_{in}	energy consumed (food digested in humans)
E_{ind}	energy stored in an inductor

Symbol	Definition
E_{out}	energy output
e	emissivity of an object
e^{+}	antielectron or positron
eV	electron volt
F	farad (unit of capacitance, a coulomb per volt)
F	focal point of a lens
$\mathbf{F}$	force
F	magnitude of a force
F	restoring force
F_{B}	buoyant force

Symbol	Definition
F_{c}	centripetal force
F_{i}	force input
$\mathbf{F}_{\mathrm{net}}$	net force
F_{o}	force output
FM	frequency modulation
f	focal length
f	frequency
f_0	resonant frequency of a resistance, inductance, and capacitance (RLC) series circuit
f_0	threshold frequency for a particular material (photoelectric effect)

Symbol	Definition
f_1	fundamental
f_2	first overtone
f_3	second overtone
f_{B}	beat frequency
f_{k}	magnitude of kinetic friction
f_{s}	magnitude of static friction
G	gravitational constant
G	green quark color
G	antigreen (magenta) antiquark color

Symbol	Definition
g	acceleration due to gravity
g	gluons (carrier particles for strong nuclear force)
h	change in vertical position
h	height above some reference point
h	maximum height of a projectile
h	Planck's constant
hf	photon energy
h_{i}	height of the image
h_{o}	height of the object
I	electric current

Symbol	Definition
I	intensity
I	intensity of a transmitted wave
I	moment of inertia (also called rotational inertia)
I_0	intensity of a polarized wave before passing through a filter
I_{ave}	average intensity for a continuous sinusoidal electromagnetic wave
I_{rms}	average current
J	joule
J/Ψ	Joules/psi meson
K	kelvin
k	Boltzmann constant

Symbol	Definition
k	force constant of a spring
K_α	x rays created when an electron falls into an $n=1$ shell vacancy from the $n=3$ shell
K_β	x rays created when an electron falls into an $n=2$ shell vacancy from the $n=3$ shell
kcal	kilocalorie
KE	translational kinetic energy
$\mathrm{KE}+\mathrm{PE}$	mechanical energy
KE_e	kinetic energy of an ejected electron
$\mathrm{KE}_{\mathrm{rel}}$	relativistic kinetic energy
$\mathrm{KE}_{\mathrm{rot}}$	rotational kinetic energy
KE	thermal energy

Symbol	Definition
kg	kilogram (a fundamental SI unit of mass)
L	angular momentum
L	liter
L	magnitude of angular momentum
L	self-inductance
ℓ	angular momentum quantum number
L_α	x rays created when an electron falls into an $n = 2$ shell from the $n = 3$ shell
L_e	electron total family number
L_μ	muon family total number
L_τ	tau family total number

Symbol	Definition
L_{f}	heat of fusion
L_{f} and L_{v}	latent heat coefficients
$\mathrm{L}_{\mathrm{orb}}$	orbital angular momentum
L_{s}	heat of sublimation
L_{v}	heat of vaporization
L_z	*z* - component of the angular momentum
M	angular magnification
M	mutual inductance
m	indicates metastable state
m	magnification

Symbol	Definition
m	mass
m	mass of an object as measured by a person at rest relative to the object
m	meter (a fundamental SI unit of length)
m	order of interference
m	overall magnification (product of the individual magnifications)
$m\left({}^{A}\mathrm{X}\right)$	atomic mass of a nuclide
MA	mechanical advantage
m_{e}	magnification of the eyepiece
m_e	mass of the electron
m_ℓ	angular momentum projection quantum number

Symbol	Definition
m_n	mass of a neutron
m_{o}	magnification of the objective lens
mol	mole
m_p	mass of a proton
m_{s}	spin projection quantum number
N	magnitude of the normal force
N	newton
$\mathbf{N}$	normal force
N	number of neutrons
n	index of refraction

Symbol	Definition
n	number of free charges per unit volume
N_{A}	Avogadro's number
N_{r}	Reynolds number
$\mathrm{N} \cdot \mathrm{m}$	newton-meter (work-energy unit)
$\mathrm{N} \cdot \mathrm{m}$	newtons times meters (SI unit of torque)
OE	other energy
P	power
P	power of a lens
P	pressure
$\mathbf{p}$	momentum

Symbol	Definition
p	momentum magnitude
p	relativistic momentum
$\mathbf{p}_{\text{tot}}$	total momentum
$\mathbf{p}'_{\text{tot}}$	total momentum some time later
P_{abs}	absolute pressure
P_{atm}	atmospheric pressure
P_{atm}	standard atmospheric pressure
PE	potential energy
PE_{el}	elastic potential energy
PE_{elec}	electric potential energy

Symbol	Definition
$\mathrm{PE_s}$	potential energy of a spring
P_g	gauge pressure
P_in	power consumption or input
P_out	useful power output going into useful work or a desired, form of energy
Q	latent heat
Q	net heat transferred into a system
Q	flow rate—volume per unit time flowing past a point
$+Q$	positive charge
$-Q$	negative charge

Symbol	Definition
q	electron charge
q_p	charge of a proton
q	test charge
QF	quality factor
R	activity, the rate of decay
R	radius of curvature of a spherical mirror
R	red quark color
R	antired (cyan) quark color
R	resistance
R	resultant or total displacement

Symbol	Definition
R	Rydberg constant
R	universal gas constant
r	distance from pivot point to the point where a force is applied
r	internal resistance
$r_{\perp}$	perpendicular lever arm
r	radius of a nucleus
r	radius of curvature
r	resistivity
r or rad	radiation dose unit
rem	roentgen equivalent man

Symbol	Definition
rad	radian
RBE	relative biological effectiveness
RC	resistor and capacitor circuit
rms	root mean square
r_n	radius of the *n*th H-atom orbit
R_{p}	total resistance of a parallel connection
R_{s}	total resistance of a series connection
R_{s}	Schwarzschild radius
S	entropy
$\mathbf{S}$	intrinsic spin (intrinsic angular momentum)

Symbol	**Definition**
S	magnitude of the intrinsic (internal) spin angular momentum
S	shear modulus
S	strangeness quantum number
s	quark flavor strange
s	second (fundamental SI unit of time)
s	spin quantum number
$\mathbf{s}$	total displacement
$\sec\theta$	secant
$\sin\theta$	sine
s_z	*z*-component of spin angular momentum

Symbol	Definition
T	period—time to complete one oscillation
T	temperature
T_{c}	critical temperature—temperature below which a material becomes a superconductor
T	tension
T	tesla (magnetic field strength *B*)
t	quark flavor top or truth
t	time
$t_{1/2}$	half-life—the time in which half of the original nuclei decay
$\tan\theta$	tangent
U	internal energy

Symbol	Definition
u	quark flavor up
u	unified atomic mass unit
$\mathbf{u}$	velocity of an object relative to an observer
$\mathbf{u}'$	velocity relative to another observer
V	electric potential
V	terminal voltage
V	volt (unit)
V	volume
$\mathbf{v}$	relative velocity between two observers
v	speed of light in a material

Symbol	Definition
$\mathbf{v}$	velocity
$\mathbf{v}$	average fluid velocity
$V_{\mathrm{B}} - V_{\mathrm{A}}$	change in potential
$\mathbf{v}_{\mathrm{d}}$	drift velocity
V_{p}	transformer input voltage
V_{rms}	rms voltage
V_{s}	transformer output voltage
$\mathbf{v}_{\mathrm{tot}}$	total velocity
v_{w}	propagation speed of sound or other wave
$\mathbf{v}_{\mathrm{w}}$	wave velocity

Symbol	Definition
W	work
W	net work done by a system
W	watt
w	weight
w_{fl}	weight of the fluid displaced by an object
W_{c}	total work done by all conservative forces
W_{nc}	total work done by all nonconservative forces
W_{out}	useful work output
X	amplitude
X	symbol for an element

Symbol	Definition
${}^{Z}_{A}X_N$	notation for a particular nuclide
x	deformation or displacement from equilibrium
x	displacement of a spring from its undeformed position
x	horizontal axis
X_{C}	capacitive reactance
X_{L}	inductive reactance
x_{rms}	root mean square diffusion distance
y	vertical axis
Y	elastic modulus or Young's modulus
Z	atomic number (number of protons in a nucleus)

Symbol	Definition
Z	impedance